The Dissertation Committee for Ross Evan Heath
certifies that this is the approved version of the following dissertation:

Analysis of the Discontinuous Galerkin Method Applied to Collisionless Plasma Physics

Committee:

Irene Gamba, Supervisor

Philip Morrison

Clint Dawson

Graham Carey

Bjorn Engquist

Analysis of the Discontinuous Galerkin Method Applied to Collisionless Plasma Physics

by

Ross Evan Heath, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2007

To my wife and sons,

Sharon, Eli and Matan.

Acknowledgments

I have the privilege of being able to thank my advisor, Irene M. Gamba. Without her direction, patience, and enduring support, this work would not have been possible. In particular, I want to thank her also for introducing me to the kinetic theory and for imparting to me her desire to use the DG methods to approximate kinetic equations. I would also like to thank Phil Morrison for introducing me to the fascinating phenomena of plasma systems, especially to Landau damping and to the traveling wave problem. I also thank Clint Dawson for his advice and expertise concerning DG methods. I also owe a debt of gratitude to my fellow ICES graduate students, both past and present, for the endless discussions that we have had, from which I have greatly benefitted.

Finally, I would like to thank my wife, Sharon. Words alone cannot convey my gratitude, appreciation, and love that I have for you. This dissertation is a product of the unconditional love and support that you have given me over the years.

Ross Evan Heath

The University of Texas at Austin
May 2007

Analysis of the Discontinuous Galerkin Method Applied to Collisionless Plasma Physics

Publication No. _____

Ross Evan Heath, Ph.D.
The University of Texas at Austin, 2007

Supervisor: Irene Gamba

Two discontinuous Galerkin methods (DG), the discontinuous flow upwind Galerkin (DFUG) and discontinuous flow upwind Galerkin-Nonsymmetric Interior Penalty Galerkin (DFUG-NIPG) methods, are proposed to approximate the Vlasov system for a perturbed flow and the Vlasov-Poisson system, respectively. These methods are chosen due to their local nature, local conservation properties, approximation properties, and their potential for hp-refinement and parallelizability.

A new optimal inverse inequality is proved for polynomials in this dissertation. Using this new inequality, an hp-optimal error estimate is proved for the NIPG method. Moreover, an error estimate is derived for the Poisson equation satisfying a Dirichlet boundary condition, where the righthandside of the equation is defined by a perturbed source term.

A new method, the DFUG method, for the Vlasov equation in six dimensional phase-space is formulated such that the method is well-defined for flows that are discontinuous across the mesh faces. Stability and $h-$optimal convergence results are proved for the method. An error estimate is proved for the error between a solution to the Vlasov system that is defined by a given flow and a solution to the Vlasov system that is defined by a perturbed flow. Explicit conditions are given as to how well a perturbed flow must approximate a given flow in order to achieve an optimal error estimate.

A new method, the DFUG-NIPG method, is proposed to approximate to the Vlasov-Poisson system in six dimensional phase-space. In the case that a discrete solution resulting from the DFUG-NIPG formulation exists, a partial hp-error estimate is proved for the error between the true solution to the Vlasov-Poisson system and the discrete solution.

DG methods are applied to three benchmark examples and a fourth experimental example. The first two benchmarks are to compute numerical solutions to the Vlasov-Poisson system that is linearized about the Maxwellian distribution, in the first example, and the Lorentzian distribution, in the second example, in order to verify that the correct Landau damping decay rates for the electric field waves are obtained up to two digit decimal accuracy. The third benchmark is to compute a numerical solution to the Vlasov-Poisson-Fokker-Planck system to check that the results correspond with currently existing results obtained using other numerical approaches. The third example is to compute a numerical solution to the Vlasov-Poisson system that is subjected to an external force field function for a fixed amount of time to determine if any BGK-like modes are present in the numerical solution.

Contents

Chapter 1

Introduction

1.1 Motivation

The Vlasov-Poisson system is a kinetic system of partial differential equations that models the time evolution of a collisonless plasma consisting of electrons and a uniform background of ions. The time-evolving quantity being modeled is the distribution function of the electrons. The Vlasov equation is used to model the transort and the acceleration of the electrons, where the acceleration is due to the self-consistent electric field and a given external field, where the electric field is the gradient of a potential that satisfies the Poisson equation defined by a righthandside forcing function that depends on the electron distribution and the supplied external field. Due to the fact that the forcing function in the Poisson equation depends on the electron distribution function, the Vlasov-Poisson system is a highly nonlinear coupled system.

Due to the nature of the Vlasov-Poisson system, many well-known phenomena for the distribution function are possible. One particular example of such a phenemonon is that of filamentation, which is due to the properties of the Vlasov equation [5]. This property causes the slopes in the velocity direction of the distribution to increase without bound as time evolves.

Other interesting phenomena exist as well as filamention, such as electric field wave-particle interactions, electron holes, ion holes, and double layers. The holes and double layers are referred to as BGK modes, which occur in plasmas that are far away from thermodynamic equilibrium [70]. Moreover, these holes are known to be very complicated vortical phase space structures that can travel quickly in time.

In 1946, Landau proved a now famous wave-particle interaction effect,known as Landau damping, that occurs in an unmagnetized, collisionless plasma [59]. Landau mathematically showed that if the one dimensional Vlasov-Poisson system is linearized about a Maxwellian distribution, or certain other equilibrium distributions, with an initial small perturbtion, then the electric field waves suffer a damping effect. This damping effect results from the fact that the wave phase-velocity is greater than the velocities of the particles. Therefore, the wave transfers a net energy to the particles, and as a result, the wave loses energy and is damped in time. Moreover, Landau's original work gave analytic expressions that can be evaluated to give the rate at which the damping occurs in the time-asymptotic limit for the most dominant mode of the electric field. Landau's work is a significant achievement in plasma physics, as it demonstrated that there is a dissipation mechanism present in collisonless plasmas.

An example of a physical process that can give rise to BGK modes is the plasma-laser interaction system that results when a laser beam propagates through a gas-filled target, as is done in fusion reactions. The electrostatic force of the laser beam, known as a ponderomotive force, results in the increase of the thermal energy of the target system, which causes ionization of the molecules within the system. This results in a phase transition from a gaseous state to that of a plasma state. Continued exposure of the plasma to the laser beam can lead to the formation of coherent structures in the phase space of the plasma. Some particular structures of interest are kinetic electrostatic electron nonlinear (KEEN) waves, which are nonlinear, nonstationary, stable, and long-lived waves in the phase space of a plasma system. Whether or not KEEN waves come into existence in a plasma exposed to a ponderomotive force is currently an unanswered question [3].

Another example of a physical process that involves a laser-plasma interaction that can give rise to various wave structures is the laser-driven inertial confinement fusion (ICF). During this process, laser beams are directed into a fusion chamber that contains a target surrounded by a hydrogen fuel. The laser heats the target to a point such that a thermonuclear burn results within the target, which leads to a massive energy release in the form of heat. However, during this process, consideration must be given to the interaction between the laser and the hydrogen fuel surrounding the target within the chamber, since the behavior of this fuel can adversely affect the desired release of heat energy by the target. Thus, understanding the types of wave structures that may arise within the hydrogen plasma is important to designing an efficient ICF.

Plasma sheaths are the boundary layers formed between a plasma and a boundary surface, where the electron and ion densities are different. The thickness of the plasma sheath

is known to be of the order of a quantity known as the Debye length. The characteristic behavior of a plasma that gives rise to a boundary sheath is known as Debye shielding. This shielding refers to a plasma's ability to shield out electric potentials that are applied to it. The plasma sheath problem is almost ubiquitous in space technology, since the enviroment of any satellite orbiting the earth is a plasma. On the boundary surfaces of these satellites and any of their antennae a plasma sheath boundary layer exists.

Due to the rich applications and fascinating dynamics that can be found in plasma systems, and collisionless plasmas in particular, there is a great interest among the computational physics and the applied mathematics communities in developing, analyzing, and implementing numerical methods to approximate the Vlasov-Poisson system. However, the complexities present in collisionless plasmas that make them so appealing also make approximating these systems with a known degree of accuracy difficult. Compounding the problem is that many realistic applications of interest involve evolving a given plasma over very long time scales. In cases such as these, small errors at each time step can accumulate as time grows, so that the final results are meaningless. However, as difficult as the task of developing numerical methods for plasma systems is, there are also many advantages in taking on this challenge.

The main advantage in developing numerical methods for plasma systems, and the Vlasov-Poisson system in particular, comes from the fact that there are many established mathematical and physical results for these systems. Properly utilized, these results can serve as a means for checking the accuracy of a numerical scheme, both in terms of analytical and computed results. For example, computational results can be used to see if they are reasonably satisfying the well-known conservation laws for the Vlasov-Poisson system. From an analysis viewpoint, one could try to establish that the numerical method under consideration yields discrete solutions that satisfy similar conservation laws as the true distribution does. Perturbation problems about equilibrium distributions, such as the Maxwellian and Lorenztian distributions, that produce a damping effect in the electric field wave, where the long-time limit of the decay rate is known, provide tangible benchmarks by which the accuracy of a numerical method can be checked.

The discontinuous Galerkin (DG) method is a finite element method that has received much attention in the last few years, especially in solving fluid flow problems. The computational interest in the method is well warranted, as the method is able to handle globally rough solutions, has local mass conservation properties, allows for weak imposition of boundary conditions, is well-suited for hp-adaptive refinements (i.e., h-adaptivity refers to refining the mesh size and p-adaptivity refers to refining the degree of the approximating basis

functions), has less numerical diffusion than most convention algorithms, and leads to block diagonal matrices in time-dependent problems that are easily invertible by hand. From an analysis viewpoint, the DG method generally allows one to establish strong stability and convergence results that are expressible in term of h and p, and can be proven to satisfy physically desirable local conservation laws. Moreover, the convergence results usually yield exponential convergence rates in p when the true unknown solution is smooth.

In the context of plasma problems, the DG method offers substantial gains. One of the main reasons for this is that the true domain of the Vlasov-Poisson system is a subset of six dimensional space (three dimensions in x and three dimensions in v). Thus, to approximate such a system is extremely computationally expensive, both in terms of CPU time and memory. The parallelizability of the DG method in this setting would then become very important, since it would potentially allow for a large number of degrees of freedom to be used to help ensure the accuracy of the computed solution. The ability to weakly impose boundary conditions in a natural way and to employ adaptive refinement strategies would allow one to resolve the boundary layer regions where plasma sheaths are formed. The adaptivity would also allow for very local mesh and polynomial refinement in the Landau damping and laser-plasma problems in those regions of the domain where the electron and ion distributions experience large variations and require a significant number of degrees of freedom to accurately capture their behavior.

The many features of the DG method make it an ideal method for use in the approximation of collisonless plasma systems. The method can be used to approximate both the convective nature of problem that is due to the Vlasov equation and the diffusive nature of the problem that is due to the Poisson equation. Since the DG method has been rarely used in the plasma setting, a thorough mathematical analysis of the method applied to the Vlasov-Poisson system is needed to establish that it is a viable numerical approach for approximating this system. In this work, mathematical analysis and a number of basic computational experiments will be carried out for the approximation of the Poisson, Vlasov, and Vlasov-Poisson systems by DG methods.

1.2 Literature review

1.2.1 Computational plasma physics

Many of the numerical techniques for solving the Vlasov-Poisson system can be divided into two groups: those the approximate the system in the $(x, v)-$phase space directly

[58],[72],[27],[54],[36], [37],[45],[46],[50], [49],[3] and those that transform the system into a different coordinate space [5],[71],[55],[56],[57].

Those numerical approaches that treat the phase space directly do not, however, usually involve working with the Vlasov equation directly. Rather most of these types of methods take advantage of the characteristic structure of the Vlasov equation, which implies that the distribution function evolves in time along trajectories that satisfy a given ordinary differential equation system. In this case, a numerical solution is gotten by considering a finite number of particles in some initial state, and then using the characteristic ODE system to update the configuration state of the particles at every time step. Once the configuration state has been updated, this information is then used as input into a numerical routine for approximating Poisson's equation for the potential, which is then fed again as input into the characterisctic ODE system to update the configuration space. The most famous of these particle-characteristic methods is the Particle-in-Cell (PIC) method [17],[52],[39], which dates back to the late 1950's.

The PIC method has been actively developed since its inception. In 1984, Cottet and Raviart presented a concise proof of the convergence of the method for the one dimensional Vlasov-Poisson system [36], and these same authors extended their results two years later in [37]. The generalization of their method of proof to the three dimensional Vlasov-Poisson system was first given by Victory, et al., in [45], for equally spaced initial data points. This work was further extended in [46],[50],[49]. Some more recent papers on the convergence of the particle methods are given in [77], [78], [79]. The particle method seems to give reasonable results, especially in cases where the tail of the distribution is negligible and a large number of particles are not neccessary.

Methods based on the discretization of the phase-space have been proposed [72],[57],[73] and seem to more efficient than particle methods in the cases mentioned above when the particle methods do not perform well. However, these approaches seem to perform well in simple geometries of the physical space, but not in more complicated geometries.

Using the finite volume method to approximate the Vlasov-Poisson system has been investigated in [19],[27], [43], which is known to be a satisfactory method for the discretization of conservation laws. In [27], Cheng and Knorr verified that the finite volume method that they proposed captured the correct Landau damping rate, up to a time of 35, for the one dimensional Vlasov-Poisson system linearized about the Maxwellian distribution. In 2001, Filbet proved the convergence of the finite volume method for the one dimensional Vlasov-Poisson system with periodic boundary conditions.

In 2001, Zhou, Guo, and Shu performed a numerical study of Landau damping [80] using a high-order accurate hybrid spectral and finite difference scheme. They observed the expected damping effect of the electric field for the one dimensional Vlasov-Poisson systems linearized about both the Maxwellian distribution and the a double-hump distribution with sufficient decay in the velocity direction. Through their investigation, the authors observed that longer the spatial periods resulted in slower decay rates for the maximum of the electric field. One peculiar aspect of their work is that it did not compare the computed results with the known theoretical decay rates, for those cases where the decay rate can be explicitly found. This is an important benchmark, since numerical damping is oftern observed, but what is most important is to verify that the correct damping rate is achieved.

Other methods have been proposed for the Vlasov-Poisson system and for many other systems similar to the Vlasov-Poisson system, except that these systems include non-zero collison operators. A few examples of numerical methods that have been used to approximate collisonal plasmas can be found in [1], [22],[41],[44], which is by no means a comprehensive list. However, the dynamics of collisional plasma systems are extremely different than those for collisionless plasmas, so that a numerical method designed to approximate a collisional plasma is not neccessarily well suited to approximate a collisionless plasma, and vice-versa.

1.2.2 Discontinuous Galerkin method

The discontinous Galerkin (DG) method is a finite element method that has been used to approximate hyperbolic, elliptic, and parabolic partial differential equations whose solutions satisfy given initial data, in the case of time-dependent problems, and boundary data. The analysis and efficient implementation of DG methods for a variety of physically motivated problems remains an active area of research among the mathematical modeling community. The strong interest of late in these methods has resulted in their being developed at a rapid pace, both in terms of their mathematical theory and their practical applications.

The intial development of the DG methods for hyperbolic equations and for elliptic and parabolic equations occured independently, but nearly simultaneously. One of the first DG schemes for approximating the solutions to second-order elliptic equations was introduced in 1971 by Nitsche [63]. In this work, the idea of enforcing Dirichlet boundary conditions weakly, instead of strongly, through the use of a penalty term was introduced. Shortly thereafter, applications of the penalty method to Laplace's equation were proposed by Babŭska et al. in [10],[11],[14]. The method developed in these works was referred to as the weak element method and approximated the unknown solution by harmonic polynomials

satisfying a jump condition. In 1976, Douglas and Dupont weakly enforced the continuity of the stress in elliptic equations and the continuity of the flux in parabolic equations [40]. The use of penalty terms across interior faces as a means of enforcing interior continuity among adjacent elements was introduced by Wheeler [75] and Percell and Wheeler [76] in the Interior Penalty Galerkin (IPG or SIPG) finite element method. The SIPG method was employed by Arnold [7],[8] to approximate parabolic equations and nonlinear elliptic problems.

In 1973, the first DG scheme for linear hyperbolic equations was introduced by Reed and Hill for approximating a neutron transport equation. This work was followed by Lesaint and Raviart in 1974 [60], where *a priori* error estimates were proved for the DG method applied to two-dimensional, linear hyperbolic problems.

During the late 1970's, throughout the 1980's, and into the early 1990's, the DG methods did not get much attention. This resulted from a lack of computing power and available resources during those times. The need for computing power to implement the DG methods results from their expensive computational costs in terms of memory. It is precisely because of their local nature, which is their main attraction, that requires vast memory, since the elements in the mesh do not engage in node sharing.

By the mid-1990's, it became computationally feasible to reconsider the DG methods for approximating solutions to partial differential equations. The motivation for using the methods came from their conservation, local approximation, and global approximation properties. Moreover, the local nature of the methods makes their implementation very ideal for parallel computation using hp-refinement strategies, where h-refinement refers to refining the sizes of the local elements which partition the domain of the problem being considered and p-refinement refers to refining the number of local basis functions used on each element. The most common basis functions for the DG methods are polynomials, in which case p-refinement refers to adjusting the degree of the basis functions. The local nature of the methods also allows for the polynomials of differing degrees to be used on different elements.

In 1997, Oden, Babŭska, and Baumann (OBB) proposed a non-symmetric DG method for approximating both diffusion and convection-diffusion problems [64],[65],[16]. The nonsymmetric form of this method has a desirable cancelation effect with respect to the errors across the inter-element boundaries. Moreover, the method exhibits a local mass conservation property at the element level, which an important fact when solving convection and/or diffusion problems. Also, the OBB method does require a penalty term in order for the

method to convergence. In 2000, Riviere et al. [68] introduced the Nonsymmetric Interior Penalty Galerkin formulation to approximate Poisson's equation with a reaction term. This method is similar to the OBB method, except that NIPG adds interior penalty terms to the formulation. In [68], the NIPG method was shown to be stable, and an $h-$optimal *a priori* error estimate in the H^1-norm was proved for the case when the reaction term was bounded away from zero. When the reaction term is set identically to zero, an $h-$optimal *a priori* error estimate in the L^2-norm was proved for the gradient of the solution, but no error estimate was given for the solution itself. NIPG has the advantage over SIPG that it does not requires a sufficiently large penalty parameter to establish convergence. In fact, all that is required of the penalty parameters is that they are uniformly bounded above and below by positive constants. The Incomplete Interior Penalty Galerkin (IIPG) method was proposed in 2003 by Dawson, Shu, and Wheeler in [38] to improve the performance of the penalty methods.

In 1997, Bassi and Rebay [15] proposed a non-penalty DG formulation for the compressible Navier-Stokes equations. A generalization of their method was introduced by Cockburn and Shu [35] and extended by Dawson and Cockburn in [29]. A unified analysis of the DG methods for elliptic problems can be found in [9] and a comprehensive review of DG methods up to 2000 can be found in [32].

1.3 Major contributions

The main contributions of this dissertation are in the fields of computational plasma physics and discontinuous Galerkin methods. These contributions are listed as follows:

- An *hp*-optimal inverse inequality in the L^2-norm for tensor-product polynomial functions is proved. This result is an improvement over an analogous currently existing inequalty.
- An *hp*-optimal error estimate is proved the NIPG method. This result is an improvement over the current error estimate for the method, which is optimal in h but suboptimal in p. This error estimate is extended to the case when approximating Poisson's equation using a source function that is some arbitrary perturbation of the true source function. This result is proved under the assumption that only Dirichlet boundary condtions are enforced. The proof does not rely on an inverse inequality holding for the Laplace operator, but instead utilizes on Poincare's inequality.
- The DFUG method is introduced to approximate the Vlasov equation defined by

a discontinous flow. This method is shown to be consistent, stable, and gives an $h-$optimal error estimate with respect to the standard upwind Galerkin norm.

- An error estimate is proved for the difference of the solution to the Vlasov system and the solution to the same Vlasov system, except the flow defining this system is some perturbation of the original flow. The resulting estimate is structured in a way such that it contains an $h-$optimal error estimate arising from the DFUG discretization of the Vlasov system and it contains a seperate contribution coming from the L^2-normed difference of the original flow and the perturbed flow taken over the elements and faces of the mesh. This estimate is then used to see the find the precise approximation conditions that the perturbed flow must satisfy, with respect to the flow of interest, in order that the error between the original Vlasov system and the perturbed Vlasov system remains optimal in h.
- The DFUG-NIPG method is introduced to discretize the Vlasov-Poisson system in phase-space. An explicit *a priori* h and partial p error estimate is proved in the case that there exists a discrete solution to the DFUG-NIPG formulation of the Vlasov-Poisson system.

1.4 Dissertation outline

In Chapter 2, a brief introduction is given for the Vlasov-Poisson system. A more in-depth introduction is given concerning the finite element method, and, in particular, the DG method. The basic notations, mesh properties, mathematical interpolation results, and a few inequalities for polynomial functions are discussed. Also, a new, optimal inverse inequality for polynomials is proved in this chapter, which is an improvement over a current analogous result.

Chapter 3 is devoted to a discussion of the NIPG method. First, the formulation for the method will be derived. We will then give a detailed discussion of the existing error analysis results. The current error estimate established for the NIPG scheme will be seen to be optimal in h and is suboptimal p by a factor of $1/2$. Through this discussion of the error analysis, the terms that cause the suboptimality in the error estimate will be determined. It will be seen that the improved inverse inequality for polynomials that was proved in Chapter 2 can be used to improve the convergence order of these limiting terms. After inserting new bounds for these terms, which will derived, and them plugging them into the original NIPG error estimate, the *hp*-optimality of the error estimate will be proved. To

our knowledge, this is a new result. We end Chapter 3 by deriving an error estimate for the NIPG method for the case when the source term is replaced by a perturbed source term. This result will be needed later on, during the discussion of the Vlasov-Poisson system.

A new method, the DFUG method, for the Vlasov equation in six dimensional phase-space is formulated such that the method is well-defined for flows that are discontinuous across the mesh. Stability and optimal convergence results are proved for the method. An error estimate is proved for the error between a solution to the Vlasov system defined by a given flow and a solution to the Vlasov system defined a perturbed flow. Explicit conditions are given as to how well a perturbed flow must approximate a given flow in order to achieve an optimal error estimate between the true solution to the Vlasov sytem and the solution to the Vlasov system defined by the perturbed flow.

A new method, the DFUG-NIPG method, is proposed to approximate to the Vlasov-Poisson system in six dimensional phase-space. In the case that a discrete solution resulting from the DFUG-NIPG formulation exists, a partial hp-error estimate is proved for the error between the true solution to the Vlasov-Poisson system and the discrete solution.

In Chapter 5, three numerical examples will investigated. Computed solutions for each of the examples will be presented. The computational results of the first two examples will serve as benchmarks for the accuracy of the DFUG-NIPG method for approximating Vlasov-Poisson examples. The third example is an experimental problem, where the theoretical results of the problem remain unclear.

Lastly, the major conclusions of this study are summarized in Chapter 6. Also, the directions for our future work will be discussed.

Chapter 2

Preliminaries and notation

2.1 Kinetic theory

The problem of describing fluid flow has been of interest to engineers and scientists for quite some time. Much of the attention has focused on fluid descriptions based upon hydrodynamical models. An implicit assumption of these models is that the velocity distribution of the species under consideration is Gaussian everywhere. For many problems in the applied sciences, this assumption can be justified and the use these models leads to sufficiently accurate results. However, there are many observed phenomena, especially in rarified gas and plasma dynamics, that arise from non-Gaussian velocity distributions for the species under consideration. In cases such as these, models based upon the kinetic theory are preferred, since they more accurately model the non-Gaussian behavior of the velocity distributions in question.

The kinetic theory is developed by first looking at a given particle system from a microscopic point of view and then using a series of arguments, that are credited in part to James Clerk Maxwell and ultimately to Ludwig Eduard Boltzmann, to arrive at at mesoscopic description of the system [25]. The system is assumed to be comprised of individual particles that are in continuous motion, where the motion has both random and non-random components. The degree of the nonrandomness of the system depends on the amount of movement at the macroscopic level. If at the macroscopic level no movement is present, then the motion is purely random and the particles are undergoing constant collisions with each other and with the boundary walls, if the system is contained. During these collisions, the directions and magnitudes of the velocities of colliding particles change in a discontinuous manner,

more or less.

In most applications, the bulk behavior of a given particle system is the main interest. Specifically, finding observable quantities such as the density, mean velocity, pressure, temperature, and viscosity are desired. Thus, the goal of kinetic theory is to start with a billiard-ball model to derive characteristics using this model, and then to use these characteristics to understand the macroscopic properties of the given system.

2.1.1 General form of kinetic systems

We now give a very basic introduction to kinetic systems of equations. For a more thorough treatment of the material presented here, the reader should consult [24],[25]. Introductory presentations from a plasma perspective can be found in [18],[47], [62],[28],[26]. We remark that throughout this work only unmagnetized plasmas will be considered.

The kinetic equations arising from the kinetic theory are partial differential or integro-partial differential equations where the unknown function, f, satisfying the given equations can be thought of as a probability distribution function (pdf), possibly upon a rescaling of this quantity. In a multispecies system, such as is the case of a plasma in which the ions are not assumed to be stationary on the time-scale being considered, the distribution of species α is denoted by f_α. The independent variables of f are time, space, and velocity, hence $f = f(\,t, x, v\,)$. Thus, for three dimensional physical space, f is a function of seven independent variables.

It is important to note that f can be interpreted as a mass distribution function, which results from scaling the pdf by the total number of particles in the system. Without loss of generality, f will always be assumed to be a pdf. Hence, the quantity

$$f_\alpha(x, v, t)\; dx_1 dx_2 dx_3 dv_1 dv_2 dv_3$$

is the probability of finding a particle, at time t, of species α in the 6-d volume element $dx_1\, dx_2\, dx_3\, dv_1\, dv_2\, dv_3$, centered at the point (x, v).

There are many types of kinetic equations. However, they all share a general framework. The equations are usually comprised of an advective transport term

$$v \cdot \nabla_x f \;, \tag{2.1}$$

a macroscopic force field term

$$\pm\, E(x,t)\cdot\nabla_v f\ , \tag{2.2}$$

and a particle collision or interaction term usually denoted by $Q(f)$. Throughout the rest of this work, we always assume that the sign of the force field E is -1. Thus, the general form of a kinetic equation is

$$f_t(x,v,t)\ +\ v\cdot\nabla_x f(x,v,t)\ -\ E(x,t)\cdot\nabla_v f(x,v,t)\ =\ Q(f)(x,v,t), \tag{2.3}$$

for $t\in(0,T)$, for some fixed $T>0$, $\quad x\in\Omega^x\subseteq\mathbb{R}^3$ and $v\in\mathbb{R}^3$, subject to some given initial condition

$$f(\,t=0,x,v\,)=f_0(\,x,v\,)\,, \tag{2.4}$$

and to appropiate boundary conditions, where Ω^x is the spatial domain of the particle system under consideration. The differential operator defining the lefthandside of (2.3),i.e.,

$$\frac{\partial}{\partial t}+v\cdot\nabla_x-E\cdot\nabla_v$$

is the well-known Vlasov operator, where the flow vector α to be

$$\alpha(x,v,t)\ =\ \begin{pmatrix} v \\ -E(x,t)\end{pmatrix}. \tag{2.5}$$

The boundary conditions that must be specified depend on Ω^x. If $\Omega^x=\mathbb{R}^3$, then the conditions at infinity, $f\to 0$, as $|x|\to\infty$, for all fixed t and v, and $f\to 0$, as $|v|\to\infty$, for all fixed t and x, are imposed. For the case when Ω^x is bounded in $\mathbb{R}^3$, an appropiate boundary condition must be supplied on some subset of $\partial\Omega^x\times\mathbb{R}^3$, where this subset depends on the collision operator being used. However, due to nature of the Vlasov operator, this subset must always include the so-called spatial inflow boundary. In order to define the inflow boundary, let $\nu(x,v)=(\,\nu^x(x),\nu^v(v)\,)\in\mathbb{R}^3\times\mathbb{R}^3$ denote the outward unit normal vector to $\partial\Omega^x\times\mathbb{R}^3$. Then the inflow boundary set $\Gamma_{_I}\subseteq\partial\Omega^x\times\mathbb{R}^3$ is defined by

$$\Gamma_{_I}\ =\ \{\ (x,v)\in\partial\Omega^x\times\mathbb{R}^3:\ v\cdot\nu^x(x)\ <\ 0\ \}\,. \tag{2.6}$$

The boundary condition specified on $\Gamma_{_I}$ must be compatible with the decay condition at infinity with respect to the velocity variable.

The phase-space domain Ω is defined to be the Cartesian product of the spatial domain

Ω^x and the velocity domain $\mathbb{R}^3$, i.e., $\Omega = \Omega^x \times \mathbb{R}^3$. The ability to decompose Ω as such a product makes physical sense, since the spatial domain is determined by the geometry of the system under consideration and the velocity domain has to do with the range of velocities the particles in the system might attain. Although the likelihood of finding particles inthe system having velocities of extreme absolute values is negligible, it is most convenient to set the velocity domain $\Omega^v = \mathbb{R}^3$.

Kinetic systems are used to model phenomena arising from problems in plasma physics, granular media, semiconductor devices, and astrophysics, which is just to name a few. One of the distinguishing characteristics of various kinetic systems is the exact form of the collision operator Q chosen. There exists a vast array of possible choices for this operator, each of which models particle collisions within the system in a particular way. Among the choices are operators that capture collisionless systems, grazing collision systems, and billiard-like collision systems. The other distinguishing factor among kinetic systems is the exact nature of how the macroscopic force field E enters the kinetic system. The field E is a vector-valued function and may or may not be coupled to the unknown function f through an auxiliary equation. In the coupled case, the field is also an unknown function. The specification of the collision and force field terms in any given kinetic system is determined by the particular problem being considered. Often times, numerical considerations will motivate these decisions, since some choices may make the numerics rather cumbersome.

Macroscopic Quantities

As mentioned already, the kinetic theory explains macroscopic quantities using mesoscopic quantities. In particular, the pdf f, which is a mesoscopic quantity, is used to define many physically relevant macroscopic quantities. In the charged particle setting, some of these quantities are the electron density, $\rho = \rho(x,t)$, which defined by integrating f with respect to v:

$$\rho(x,t) := \int_{\mathbb{R}^N} f(x,v,t)\,dv.$$

The mean current $j(x,t)$ of the charged particle system is the average of the velocities at time t and position x:

$$j(x,t) := \int_{\mathbb{R}^N} v f(x,v,t)\,dv.$$

For a system in equilibrium, the mean current is identically equal to zero. Other useful quantities are the momentum flow matrix, $m = (m_{i,j})_{i,j=1}^{N}$, defined as:

$$m_{i,j}(x,t) := \int_{\mathbb{R}^N} v_i v_j f(x,v,t)\, dv;$$

the energy density per unit volume, $w(x,t)$, defined as:

$$w(x,t) := \frac{1}{2} \int_{\mathbb{R}^N} |v|^2 f(x,v,t)\, dv;$$

the energy flow vector, $r = (r_i)_{i=1}^{N}$, defined by:

$$r_i(x,t) := \frac{1}{2} \int_{\mathbb{R}^N} v_i |v|^2 f(x,v,t)\, dv.$$

The above macroscopic quantities are easily defined, assuming that the pdf f is known. In many special cases, one can analytically find f by solving (2.3) or (2.7) using various methods. However, in most situations, solving for f by analytical means is an extremely difficult task, if not impossible. Thus, in most cases, one must employ numerical techniques to find an approximate solution to f.

Force Field Coupling

A common coupling for the force field is the Poisson coupling. This coupling is used when in plasma systems to model the Coulombic forces acting between charged particles that are proportional to the inverse of the squared distance between them. It should be noted that this coupling, which is the only coupling considered in this work, ignores the contribution of the magnetic field to the overall force field. However, this coupling does allow for a fixed background force $C(x,t)$, which is called the doping force for charged particle systems. The general form of a Poisson-coupled kinetic system equations is

$$\begin{aligned} f_t(x,v,t) \;+\; v \cdot \nabla_x f(x,v,t) \;-\; E(x,t) \cdot \nabla_v f(x,v,t) &= Q(f)(x,v,t), && (2.7)\\ E(x,t) &= -\nabla_x \psi(x,t), && (2.8)\\ \nabla_x \cdot (\varepsilon(x) \nabla_x \psi(x,t)) &= \rho(f)(x,t) - C(x,t)\,, && (2.9) \end{aligned}$$

for $t > 0,\ x \in \Omega^x \subseteq \mathbb{R}^3$, and $v \in \mathbb{R}^3$, subject to the initial condition

$$f(t = 0, x, v) = f_0(x, v) \tag{2.10}$$

and to appropiate boundary conditions if Ω^x is bounded in $\mathbb{R}^3$. Here $\varepsilon(x)$ is a given function and the function $\rho(f)(x, t)$ is the density, which was previously defined.

So far, we have introduced two types of kinetic systems of equations: the uncoupled kinetic system (1.1)-(1.2) and the Poisson-coupled kinetic system (1.4)-(1.7). Of course, if Ω^x were bounded, then both the uncoupled and the coupled systems would have to be augmented by at least two more equations, which would specify the spatial boundary conditions to be satisfied by f and ψ. Further discussion of boundary conditions will be held off until the Fokker-Planck collision operator is introduced.

We see from looking at the Poisson-coupled kinetic system, that this system is a nonlinear system due to the nature of the Poisson coupling. As for the uncoupled system, the system is linear if Q is linear and is nonlinear if Q is nonlinear.

Collision Operators

Although the focus of this work concerns collisionless plasma systems, it is important to understand the role that collision operators play in kinetic systems. Only by being aware of the role collision operators play in driving particle systems to steady-state solutions can we appreciate the complicated dynamics that may arise in collisionless systems, where there is not, in general, any dissipative mechanism that causes the system to converge to a steady-state solution.

The most famous collision operator is the Boltzmann collision operator, which assumes that the only type of collisions particles undergo are binary elastic (billiard-like) collisions. However, there exists many variants of the Boltzmann operator and many other non-Boltzmann operators. For a thorough introduction to the Boltzmann equation, the interested reader should consult [24],[25].

Although there are many different collision operators, physical considerations lead to common properties that each operator should possess. Some common properties of nonzero collision operators are that they

- preserve mass ;
- act only on the velocity variable v ;

- are dissipative in the sense that collisions cause a particular entropy functional to increase, and this functional is maximized for some subclass of Gaussian distributions.

We note that the above dissipative property of collision operators certainly does not hold in the collisionless case when $Q \equiv 0$.

The mass preservation property means that mass is neither created or destroyed. Any change in the mass of the system must result from particles entering or exiting the system at the boundaries of the spatial domain. If the spatial domain $\Omega^x = \mathbb{R}^3$, then the total mass of the system does not vary with time and it is equal to $\int_{\Omega^x} \int_{\mathbb{R}^3} f_0(x, v)\, dxdv$. When we consider f to be a pdf, then the total mass is of course one. It should be noted that the total mass of a system having a bounded spatial domain can remain constant over time if certain boundary conditions are imposed, with periodic boundary conditions being the most obvious.

The dissipative property comes from the fact that many of the physical systems described by the kinetic theory tend to converge to an equilibrium state in the asymptotic time-limit, assuming there are no external forces disturbing the system.

Dissipation is usually expressed mathematically through a particular functional $H(f)(t)$, known as the entropy functional. This functional is used in determining the convergence rate of the distribution function $f(x, v, t)$, which is the solution to some given kinetic system, to its equilibrium state $f_\infty(x, v)$. A basic outline for establishing these rates using the entropic method can be found in [6] and roughly goes as follows:

- Identify the equilibrium state f_∞ and the entropy functional H for the given kinetic system [74],[51]. H is chosen so that it attains its maximum value at f_∞.
- Define the relative entropy functional $H(f|f_\infty) = H(f) - H(f_\infty)$, which is used for measuring the distance, in an entropy sense, between f and f_∞.
- Define the entropy production functional $I(f)(t) = -\frac{d}{dt}H(f)(t)$.
- Prove a functional inequality of the form $I(f) \geq \theta(H(f|f_\infty))$, where θ is some continuous function, that is strictly positive when $H > 0$.

The function θ in the above functional inequality determines the exact nature of the decay rate. To see this, we first note that the functional inequality can be equivalently written as $-\frac{d}{dt}H(f|f_\infty) \geq \theta(H(f|f_\infty))$. One way to use this inequality to establish a decay rate for the relative entropy is to bound some positive power of $H(f|f_\infty)$ by $\theta(H(f|f_\infty))$. If such a bound is obtained, then ordinary differential equation (ODE) techniques can be used to

give a convergence rate.

The two most common bounds are a linear bound and a higher-degree polynomial bound. These bounds take the form $\theta(H) \geq \lambda H$ and $\theta(H) \geq \lambda H^{1+\alpha}$, for $\alpha \geq 1$, respectively. For the linear bound, the inequality $-\frac{d}{dt}H(f|f_\infty) \geq \lambda H(f|f_\infty)$ is obtained. This implies that $H(f|f_\infty)(t) \leq H(f|f_\infty)(0)e^{-\lambda t}$. For the higher-degree polynomial bound, the inequality $-\frac{d}{dt}H(f|f_\infty) \geq \lambda H(f|f_\infty)^{1+\alpha}$ holds. This leads to the bound $H(f|f_\infty)(t) \leq (\frac{H^{-\alpha}(f|f_\infty)(0)}{t} - \alpha\lambda)t^{-\frac{1}{\alpha}}$. Since $\frac{H^{-\alpha}(f|f_\infty)(0)}{t}$ tends to 0 as $t \to 0$, it follows that the relative entropy decays like $t^{-\frac{1}{\alpha}}$. Thus, the case of the linear bound results in the exponential decay of the relative entropy, whereas the case of the higher-degree polynomial bound results in the algebraic decay of the relative entropy.

The entropic method gives convergence rates for the relative entropy. To establish convergence rates with respect to the L^1-norm, i.e, $\|f(t)-f_\infty\|_{L^1(\Omega^x\times\mathbb{R}^3)}$, one must bound the norm by the relative entropy. A well-known example of such a bound that holds for Fokker-Planck type equations is the Csiszar-Kullback-Pinsker inequality, which is $\|f(t) - f_\infty\|^2_{L^1(\Omega^x\times\mathbb{R}^3)} \leq CH(f|f_\infty(t))$, where C is some constant.

2.1.2 Vlasov-Poisson system of equations

The Vlasov-Poisson system of equations results when the collision operator is taken to be identically zero. These equations model a system in which collisions among the particles are sufficiently rare, so that they can me completely ignored. In this case, if the magnetic field is ignored, the motion of the particles is governed according to the acceleration resulting from the electrostatic force field $E(x,t)$.

The existence and uniqueness properties of the Vlasov-Poisson system in six dimensional phase space is a current research topic. Various notions of solutions, i.e., classical, weak, mild, etc., have been proposed and investigated. Until recently, most studies of this system were done in the case that the spatial domain was unbounded [61],[66],[69]. However, recent results have extended previous results for the unbounded spatial domain to the case when the spatial domain is assumed to be a smooth, bounded, convex domain in $\mathbb{R}^3$, for absorbing or specularly reflecting boundary conditions [53].

The notion of a solution used in this work will be that of a classical solution, which will be defined shortly. The material presented in this section will follow the review given by Rein in [67], where in-depth results are summarized for the regularity of classical solutions to the Vlasov-Poisson system when the spatial domain is unbounded. Our discussion given on

the Vlasov-Poisson system when the spatial domain is a smooth, bounded, convex domain follows the original work of Hwang [53].

The Vlasov-Poisson system of equations for the full phase-space $\mathbb{R}^3 \times \mathbb{R}^3$, subject to an initial condition $f_0(x,v) \in C_c^1(\mathbb{R}^3 \times \mathbb{R}^3),\ f_0 \geq 0$, is stated as follows:

$$\begin{aligned} f_t \ + \ \alpha \cdot \nabla f &= 0\,, & \mathbb{R}^3 \times \mathbb{R}^3 \times (0,T] \quad (2.11)\\ E &= -\nabla_x \psi\,, & \mathbb{R}^3 \times (0,T] \quad (2.12)\\ -\Delta_x \psi &= \rho(f)\,, & \mathbb{R}^3 \times (0,T] \quad (2.13)\\ f(\,t=0\,) &= f_0\,, & \mathbb{R}^3 \times \mathbb{R}^3\,. \quad (2.14) \end{aligned}$$

Using the above equations, we now give the definition of a classical solution to the Vlasov-Poisson system.

Definition 1. [Classical Solution of Vlasov Poisson System for Unbounded Domain]
A function pair $(\,f,\psi\,) : (\,\mathbb{R}^3 \times \mathbb{R}^3 \times [0,T]\,) \times (\,\mathbb{R}^3 \times [0,T]\,) \rightarrow [0,\infty)$ *is a classical solution up to time* T *of the Vlasov-Poisson system (2.11)-(2.14) if* f *is continuously differentiable with respect to all of its variables, the induced density* $\rho = \int_{\mathbb{R}^3} f\,dv$ *is continuously differentiable,* ψ *is twice continuously differentiable with respect to* x *and once continuously differentiable with respect to* t*, for every compact subinterval* $J \subset [0,T]$ *the field* $\nabla_x \psi$ *is bounded on* $\mathbb{R}^3 \times J$*, and equations (2.11)-(2.14) are satisfied.*

The existence and uniqueness of global classical solutions to the Vlasov-Poisson system was proved in 1989, when two independent proofs were given [61], [66]. Shortly thereafter, another proof was given in [69]. The results in all of these proofs hold for both the attractive and repulsive cases, i.e., the results hold for both $\pm E$ in the Vlasov equation (2.11). These results are stated in the following theorem.

Theorem 1. *Let the initial condition* $f_0(x,v) \in C_c^1(\mathbb{R}^3 \times \mathbb{R}^3),\ f_0 \geq 0$*, be given. Then there exists a classical solution up to time* $+\infty$ *to the Vlasov-Poisson system (2.11)-(2.14).*

The Vlasov-Poisson system of equations for the case when the spatial domain Ω^x is bounded is similar to the above definition for the full phase space case, except that boundary conditions must be supplied for f on the inflow boundary and for the potential ψ on $\partial\Omega^x$. The

exact Vlasov-Poisson system investigated in [53] is the following:

$$
\begin{aligned}
f_t \;+\; \alpha \cdot \nabla f \;&=\; 0\,, && \Omega^x \times \mathbb{R}^3 \times (0,T]\,, && (2.15)\\
E \;&=\; -\nabla_x \psi\,, && \Omega^x \times (0,T]\,, && (2.16)\\
-\Delta_x \psi \;&=\; \rho(f)\,, && \Omega^x \times (0,T]\,, && (2.17)\\
f(\,t=0\,) \;&=\; f_{_0}\,, && \Omega^x \times \mathbb{R}^3\,, && (2.18)\\
f \;&=\; f_{_I}\,, && \Gamma_{_I} \times (0,T]\,, && (2.19)\\
\psi \;&=\; 0\,, && \partial\Omega^x \times (0,T]\,, && (2.20)
\end{aligned}
$$

where Ω^x is assumed to be a bounded, smooth, convex domain.

The boundary condition supplied for f must be compatible with the initial condition, since f must satisfy both of these conditions when $t = 0$. The following precise definition of compatibility is as given in [53], where the boundary condition specified for ψ is the Dirichlet condition $\psi = 0$ on $\partial\Omega^x$. To state the compatibility conditions, we first define the noflow boundary set $\Gamma_{_I}^0$ as

$$
\Gamma_{_I}^0 \;=\; \{\,(x,v) \in \partial\Omega^x \times \mathbb{R}^3 :\; v \cdot \nu^x \;=\; 0\,\}\,. \tag{2.21}
$$

With this definition, we now state the definition of data compatibility for the Vlasov-Poisson system when the spatial domain is a smooth, bounded, convex domain.

Definition 2. [Strict Data Compatibility for Bounded Spatial Domain] *Let $k \geq 1,\ 3 < p \leq +\infty$. Let $f_{_0} \in W^{k,p}(\Omega^x \times \mathbb{R}^3)$ and $f_{_I} \in W^{k,p}(\Gamma_{_I} \times [0,\infty))$ have compact support and $f_{_0}, f_{_I} \geq 0$. The data pair $(f_{_0}, f_{_I})$ is said to be compatible if*

$$
\begin{aligned}
\partial^\alpha f_{_0} \;&=\; \partial^\alpha f_{_I} && \text{on } \Gamma_{_I} \times \{t=0\}\,,\ \forall\ |\,\alpha\,| \leq k-1\ ; && (2.22)\\
f_{_I} \;&=\; 0 && \text{on } \Gamma_{_I}^0 \times [0,+\infty) && (2.23)\\
|\,\partial^\alpha f_{_I}\,| \;&\leq\; C\,|\,v \cdot \nu^x\,|^{|\,\alpha\,|} && \text{on } \Gamma_{_I} \times [0,+\infty),\ \forall\ |\,\alpha\,| \;=\; k\,, && (2.24)
\end{aligned}
$$

where α is a multi-index.

Using this definition, we can now state the existence and uniqueness result given in [53].

Theorem 2. *Let the initial data pair $(f_{_0}, f_{_I})$ be compatible according to Definition 2. Then there exists a solution pair $(\,f,\,\psi\,) \in W^{k,p}(\,\Omega^x \times \mathbb{R}^3 \times [0,+\infty)\,) \times W^{k+2,p}((\Omega^x \times [0,+\infty)\,)\,)$ satisfying the Vlasov-Poisson system (2.15)-(2.20).*

We mention that procedures for proving Theorems 1 and 2 involve constructing a sequence

of iterates $\{ f^n, \psi^n \}_{n=0}^{\infty}$, where the first iterate (f^0, ψ^0) is defined by setting $f^0 = f_0$ and then letting ψ^0 be the solution to the Poisson system defined by the source term f^0. Thus, both of the functions comprising the first iterate are constant in time. Then f^1 is defined to be the solution to the Vlasov system defined by the function $\nabla_x \psi^0$ and then ψ^1 is defined to be the solution to the Poisson system defined by the source term f^1. Thus, the functions comprising the second iterate are time-dependent. Following the above pattern, the entire sequence $\{ f^n, \psi^n \}_{n=0}^{\infty}$ is well-defined. We should point out that boundary condtions specified above for f and ψ in the definition of the Vlasov-Poisson system are the boundary data used to define each of the iterate pairs. After defininining this sequence, the goal is to then prove that this sequence converges to a unique fixed point pair (f, ψ), which is then shown to satisfy the Vlasov-Poisson system.

A few remarks about the assumed conditions in the above theorem are in order. Condition (2.22) is reasonable in that ensures that the inflow boundary condition at time zero corresponds to the initial condition, so that there is no mismatch on the inflow boundary. Condition (2.23) is somewhat limiting in that many inflow functions $f_{_I}$ do not satisfy this condition. Moreover, since this condition holds when $t = 0$, it also thereby places restrictions on the initial condition, via (2.22). The standard Maxwellian distribution is example of a function that is maximized when $v = 0$, but since $v \cdot \nu^x = 0$ when $v = 0$, (2.23) would be violated for any initial distribution remotely similar to the Maxwellian. One last point we mention is that the above Vlasov-Poisson sytem of equations specifies that the potential is identically zero on $\partial\Omega^x$, which is very restrictive condition from an applied perspective.

In general, there does not exist regularity theory for the types of Vlasov Poisson systems that are considered in Chapter 4. To begin with, the spatial domain that we will work with is $\Omega^x = [0, L_{_1}] \times [0, L_{_2}] \times [0, L_{_3}]$, where $L_{_1}$, $L_{_2}$, $L_{_3}$ are fixed constants. Although this domain is convex, it is not smooth. As for the initial and inflow conditions, we will want to work with conditions that are more general than those allowed for in Definition 2. Also, we will use more general Dirichlet data than $\psi = 0$ on $\partial\Omega^x$.

The approach that will be taken in Chapter 4 concerning the regularity of the Vlasov-Poisson systems of interest will be to assume that a unique classical solution exists to the system for compatible data pairs $(f_{_0}, f_{_I})$, where the definition of compatibility will be relaxed from the above definition.

The precise statement of the Vlasov-Poisson system of equations that will be considered in

Chapter 4 is the same as equations (2.15)-(2.20), except that (2.20) is replaced by

$$\psi \;=\; r_{_D}\,, \qquad\qquad \partial\Omega^x \times (0,T]\,, \tag{2.25}$$

where $r_{_D}$ is a given function in $L^2(\Omega^x \times [\,0,T\,])$ and $\Omega^x = [0,L_{_1}] \times [0,L_{_2}] \times [0,L_{_3}]$.

The relaxed definition of data pairs $(f_{_0}, f_{_I})$ that will be used in Chapter 4 is as follows:

Definition 3. *Given $T > 0$, the data pair $(f_{_0}, f_{_I})$ is said to be compatible up to time T if*

$$\begin{aligned} f_0 &\in C_c^1(\mathbb{R}^3 \times \mathbb{R}^3)\,, && (2.26)\\ f_{_I} &\in C_c^1(\partial\Omega^x \times \mathbb{R}^3 \times [0,T]), \qquad \text{, and if} && (2.27)\\ f_{_I}(t=0) &= f_{_0}\,, \qquad \forall\,(x,v) \in \Gamma_{_I}\,. && (2.28) \end{aligned}$$

In Chapter 4, the exact specification of the Vlasov-Poisson system and the definition of compatible data will again be given, so as to avoid any ambiguity. It is important to note that in the study of the DG methods as they apply to the Vlasov-Poisson system, it will be assumed that a unique classical solution to the system under consideration exists and satisfies certain regularity properties.

2.1.3 Vlasov-Poisson-Fokker-Planck system of equations

We mention here briefly a Vlasov-Poisson system having a Fokker-Planck (FP) collision operator, since one of the examples in Chapter 5 involves this system. In many applications, the Vlasov-Poisson-Fokker-Planck (VPFP) system is used to model systems of charged particles. Examples of this include semiconductor device modeling and plasma modeling. The Fokker-Planck collision operator models grazing collisions and friction effects in a fluid. The grazing collisions are accounted for by the term $\sigma\triangle_v f$, which is due to the fact that during a grazing collision the colliding particles have a slight change in velocity, and the friction effects are accounted for by the term $\nabla_v \cdot (\beta v f)$. The constants σ and β correspond to the thermal diffusivity and viscosity, respectively, of the fluid and satisy $\sigma > 0$ and $\beta > 0$. Thus, the FP collision term is given by $Q(f) = \nabla_v \cdot (\sigma\nabla_v f + \beta v f)$. So, the Vlasov-Poisson-Fokker-Planck system is then the following:

$$\begin{aligned} f_t(x,v,t) \;+\; \alpha\cdot\nabla f \;&=\; \nabla_v \cdot (\,\sigma\nabla_v f \;+\; \beta v\, f\,)\,,\\ E(x,t) \;&=\; -\nabla_x \psi(x,t)\,, \qquad \text{and} && (2.29)\\ -\Delta_x\psi(x,t)) \;&=\; \rho(f)(x,t)\,, && (2.30) \end{aligned}$$

for $t > 0,\ x \in \Omega^x \subseteq \mathbb{R}^3$, and $v \in \mathbb{R}^3$, subject to the initial condition

$$f(t = 0, x, v) = f_0(x, v) \tag{2.31}$$

and the appropiate boundary conditions if Ω^x is bounded in $\mathbb{R}^3$.

2.2 Family of meshes

The conventions and notations used to describe a family of meshes in this section will be used throughout the rest of this work, unless otherwise stated. The mesh considerations found herein are pursued with the goal of establishing a framework for the use of approximation functions that are discontinuous, or broken, over partitions of the domain $\Omega \subset \mathbb{R}^3$ or, as will been seen later on in the discussion of the Vlasov and Vlasov-Poisson system, $\Omega \subset \mathbb{R}^6$. For now, unless noted differently, it will be assumed that $\Omega \subset \mathbb{R}^3$ and that Ω is a polygonal domain. In all places, the constant C will be used to denote a generic constant that is independent of both the mesh being considered and the order of the underlying approximation space being used. Also, it will have different values at different places. When time dependent problems are being considered over an interval $[0, T]$, where $T > 0$ is fixed, the constant C will also be allowed to depend on T. A mathematically rigorous treatment of the subject matter presented in this chapter can be found in both [42] and [20], except for the results given in Lemma 2.67 and in Lemma 7.

In Lemma 6 and in Lemma 7, new polynomial inverse inequalities are proven. These new inequalities are a significant contribution to the literature, in that they improve upon currently existing suboptimal inverse inequalities. Moreover, the inequalities presented herein are in fact optimal, in a sense to be made precise later on, so that no improvement is possible for them. One of the implications of these results will be discussed at length in Chapter 3, where it is proven that a well established finite dimensional approximation method for second-order, elliptic boundary value problems that was thought to give suboptimal approximation results in fact yields optimal approximation properties.

2.2.1 Mesh basics

Let $\{\mathcal{T}_h\}_{h>0}$ be a family of partitions, also known as meshes, of the domain Ω. It is assumed that each mesh $\mathcal{T}_h$ is a collection of sets $\{K_1, \ldots, K_{N_h}\}$, called elements, satisfying the

properties

$$(i) \quad \overline{\Omega} \;=\; \cup_{j=1}^{N_h} K_{_j}\,, \tag{2.32}$$

$$(ii) \quad K_{_j} \text{ is compact, convex, for } j = 1, \ldots, N_h\,, \text{ and} \tag{2.33}$$

$$(iii) \quad \mathring{K}_i \cap \mathring{K}_j = \emptyset, \text{ for } i \neq j,\ i, j = 1, \ldots, N_h\,. \tag{2.34}$$

For a fixed mesh $\mathcal{T}_h$, we define for each element $K_{_j} \in \mathcal{T}_h$ its diameter by

$$h_{_j} \;=\; \text{diam}(\, K_{_j}\,) \;=\; \max_{x_1, x_2 \in K} |\, x_1 - x_2\,|_{\mathbb{R}^3}\,,$$

where $|\,\cdot\,|$ is used to denote the Euclidean vector-norm. The maximum diameter of the elements in the mesh is then defined by

$$h \;=\; \max_{j=1,\ldots,N_h} \{\, h_{_j}\,\}\,.$$

The value of h gives an indication of the overall level of refinement of the mesh being employed. Thus, we may think of $\{\mathcal{T}_h\}_{_{h>0}}$ as sequence of successively refined meshes. It will be always be assumed that any mesh being used has a refinement level at least satisfying $h < 1$.

Assuming, for a fixed mesh $\mathcal{T}_h$, that the mesh only satisfies properties (2.32)-(2.34) does not provide a discontinuous approximation method defined on such a mesh with enough structure for elegant error analysis or for an efficient computational implementation of the method. To overcome these difficulties, it will be assumed that any mesh in $\{\mathcal{T}_h\}_{_{h>0}}$ satisfies three additional properties: affineness, non-degeneracy, and quasi-uniformity. Each of these properties will now be discussed.

The first additional property to be assumed is that every element $K_{_j}$ is the image of the unit cube

$$\widehat{K} \;=\; [\,0, 1\,]^3\,,$$

called the reference element, under an affine, bijective transformation $T_{_{K_j}}$. A mesh satisfying this condition is called an affine mesh. Since each transformation $T_{_{K_j}}$ is affine, there exists a constant matrix $M_{_{K_j}} \in \mathbb{R}^{3\times 3}$, and a constant vector $b_{_{K_j}} \in \mathbb{R}^3$ such that

$$T_{_{K_j}} : \widehat{K} \ni \hat{x} \;\rightarrow\; x = M_{_{K_j}} \hat{x} + b_{_{K_j}} \in K_{_j}\,. \tag{2.35}$$

Since the mapping T_{K_j} is bijective, it follows that the Jacobian matrix M_{K_j} is invertible.

The fact that every element is the image of $\widehat{K}$ under an affine transformation implies that each element is in fact a parallelepiped. Thus, for $\Omega \subset \mathbb{R}^3$ each element has six faces and twelve edges. If a particular face of an element K_j lies in the interior of Ω, then this face will have a non-empty intersection with some of the faces of elements that are adjacent to K_j. The non-empty intersections resulting from the intersections of two faces from adjacent elements are referred to as mesh faces. From now on, when we speak of a face it will be understood that we are referring to a mesh face, as opposed to the face of an individual element. Only on the boundary of Ω will the mesh faces correspond the faces of individual elements, since the faces of elements that lie on $\partial\Omega$ do not intersect with any other faces of adjacent elements.

A face f_k is said to be an interior face if $|\, f_k \,| \neq 0$, where $|\, \cdot \,|$ denotes the usual Lesbegue area measure in $\mathbb{R}^2$, and if there exist two distinct elements K_1 and K_2 in $\mathcal{T}_h$ such that $f_k = K_1 \cap K_2$. Then we define K_{12} to be $K_1 \cup K_2$ and $\overline{h}_k$ to $\overline{h}_k = \max\{h_1, h_2\}$. A face f_k is said to be a boundary face if $f_k \subset \partial\Omega$ and if there exists an element K_j, denoted by K_{f_k}, such that f_k is a face of K_j. We denote the set of all faces of $\mathcal{T}_h$ by

$$\mathcal{F}_h \;=\; \{\, f_1, \ldots, f_{P_h}, f_{P_h+1}, \ldots, f_{M_h} \,\}\,, \tag{2.36}$$

where f_k is an interior face for $k = 1, \ldots, P_h$, and f_k is a boundary face for $k = P_h + 1, \ldots, M_h$. The set of all interior faces $\mathring{\mathcal{F}}_h$ is then given by

$$\mathring{\mathcal{F}}_h \;=\; \{\, f_1, \ldots, f_{P_h} \,\}\,. \tag{2.37}$$

For each $f_k \in \mathcal{F}_h$ we associate a unit normal vector ν_k. For $k > P_h$, ν_k is taken to be the outward unit normal vector to $\partial\Omega$. For $1 \leq k \leq P_h$, we fix ν_k to be one of the two unit normal vectors to f_k. We note that in the future, ν_{K_j} will be used to denote the outward unit normal vector to ∂K_j. For an interior face $f_k = K_1 \cap K_2$, it will always be assumed that K_1 is the element such that $\nu_{K_1} = \nu_k$ on f_k, which then implies that $\nu_{K_2} = -\nu_k$ on f_k.

The next property we require is that for the family of meshes $\{\mathcal{T}_h\}_{h>0}$ there exists a constant $m_0 > 0$ such that

$$\forall\, h,\ \forall\, K_j \in \mathcal{T}_h, \qquad h_j \leq m_0\, \rho_j\,, \tag{2.38}$$

where ρ_j is defined as the diameter of the largest ball that can be inscribed in K_j. A family of meshes satisfying this property is said to be nondegenerate, or regular. A nondegenerate

family of meshes $\{\mathcal{T}_h\}_{h>0}$ has the property that the elements are prevented from "collapsing", i.e., for a given element, the angle at which any two intersecting faces of the element are joined is bounded below by a constant greater than zero, as the refinement level h decreases towards zero.

The last mesh property that is assumed is quasi-uniformity. A family of meshes $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform if there is a constant $\tau > 0$ such that

$$\forall\, h,\ \forall\, K_j \in \mathcal{T}_h, \qquad h_j \geq \tau\, h\,. \tag{2.39}$$

Thus, uniformity taken together with nondegeneracy places uniform upper and lower bounds on the ratio between the volumes of any two elements of any particular mesh.

The many implications of the assumed mesh properties will be become more apparent in the remainder of this work. For now, we state two important implications that result from $\{\mathcal{T}_h\}_{h>0}$ satisfying (2.32)-(2.39). The first lemma establishes some basic facts about the Jacobian matrix M_j of a given element in the mesh.

Lemma 1. *Let $\mathcal{T}_h \in \{\mathcal{T}_h\}_{h>0}$. If $K_j \in \mathcal{T}_h$ is arbitrary, then*

$$|\, det(M_{K_j})\,| \;=\; \frac{|\,K_j\,|}{|\,\widehat{K}\,|}\,, \quad \|\,M_{K_j}\,\|_{\mathbb{R}^{3\times 3}} \;\leq\; \frac{h_j}{\rho_{\widehat{K}}}\,, \quad \text{and} \quad \|\,M_{K_j}^{-1}\,\|_{\mathbb{R}^{3\times 3}} \;\leq\; \frac{h_{\widehat{K}}}{\rho_j}\,, \tag{2.40}$$

where $\rho_{\widehat{K}}$ is the largest ball that can be inscribed in $\widehat{K}$, $h_{\widehat{K}} = \mathrm{diam}(\widehat{K})$, and $|\cdot|$ is the Lesbegue volume measure in $\mathbb{R}^3$. In particular, since $\widehat{K} = [\,0,1\,]^3$, it follows that $|\,\widehat{K}\,| = 1$, $\rho_{\widehat{K}} = \frac{\pi}{6}$ and $h_{\widehat{K}} = \sqrt{3}$.

The proof of this result can be found in [42]. The next lemma shows that the volume and the diameter of a given element K_j are intimately related.

Lemma 2. *Let $K_j \in \mathcal{T}_h$ be arbitrary. Let f_{K_j} be an arbitrary face of K_j and let e_{K_j} be an arbitrary edge of K_j. Then*

$$|\,K_j\,| \;\sim\; h_j^3 \;\sim\; h^3\,, \qquad |\,f_{K_j}\,| \;\sim\; h_j^2 \;\sim\; h^2\,, \qquad |\,e_{K_j}\,| \;\sim\; h_j \;\sim\; h\,. \tag{2.41}$$

Proof. Since the mesh is affine with respect to $\widehat{K} = [\,0,1\,]^3$, it easily follows that $|\,K_j\,| \leq h_j^3$. Now we show that K_j is uniformly bounded below by h_j^3. Since ρ_j is the diamenter of the largest ball that can be inscribed in K_j, $\frac{4\pi}{3}(\frac{\rho_j}{2})^3 \leq |\,K_j\,|$. From the nondegeneracy of the

mesh, this inequality implies

$$\frac{4\pi}{3}\left(\frac{h_j}{2m_0}\right)^3 \;\leq\; |\,K_j\,|\,. \tag{2.42}$$

Combining the above upper and lower bounds for $|\,K_j\,|$ results in

$$\frac{\pi}{6m_0^3}\,h_j^3 \;\leq\; |\,K_j\,| \;\leq\; h_j^3\,. \tag{2.43}$$

This establishes that $|\,K_j\,| \sim h_j^3$.

Let f_{K_j} be an arbitrary face of the element K_j. Then $|\,f_{K_j}\,| \leq h_j^2$. To bound $|\,f_{K_j}\,|$ below by a constant times h_j^2, we use the fact that $|\,K_j\,| \leq h_j\,|\,f_{K_j}\,|$. Combining this inequality with (2.43) leads to $\frac{\pi}{6m_0^3}\,h_j^2 \leq |\,f_{K_j}\,|$. So, we end up with

$$\frac{\pi}{6m_0^3}\,h_j^2 \;\leq\; |\,f_{K_j}\,| \;\leq\; h_j^2\,. \tag{2.44}$$

Let e_{K_j} be an arbitrary edge of the element K_j. Then $|\,e_{K_j}\,| \leq h_j$. Combining the fact that $e_{K_j}\,h_j^2 \geq |\,K_j\,|$ and inequality (2.43) leads to $\frac{\pi}{6m_0^3}\,h_j \leq |\,e_{K_j}\,|$. Hence,

$$\frac{\pi}{6m_0^3}\,h_j \;\leq\; |\,e_{K_j}\,| \;\leq\; h_j\,. \tag{2.45}$$

To complete this proof, we use the quasi-uniformity of the mesh. □

2.2.2 Broken Sobolev spaces

The broken Sobolev spaces are function spaces whose member functions are Sobolev functions over each element of a given mesh, but are allowed to be discontinuous across the boundaries of the elements. Hence, these function spaces enjoy the regularity of the Sobolev spaces locally, i.e, within any $K_j \in \mathcal{T}_h$, but this regularity does not extend globally, i.e., over the whole domain Ω.

In order to introduce the broken Sobolev spaces, we first define the usual Sobolev spaces. The Sobolev space results presented below are standard and are outlined in [42]. For an in-depth study of these spaces, the reader should consult [2].

Sobolev spaces

Let X be a measurable, open set in $\mathbb{R}^d,\ d \in \mathbb{N}$, with boundary ∂X. The general definition of the Sobolev spaces is as follows.

Definition 4. *Let $1 \leq p \leq +\infty$ and $s \geq 0$ be integers. The Sobolev space $W^{s,p}(X)$ is defined as*

$$W^{s,p}(X) \ = \ \{\, \theta \in L^p(X) : \ \partial^\alpha \theta \in L^p(X),\ \forall\, |\alpha| \leq s \,\}\,, \tag{2.46}$$

where the derivatives are understood in the distributional sense. Moreover, this space is a Banach space when equipped with the norm

$$\| \, \theta \, \|_{W^{s,p}(X)} = \sum_{|\alpha| \leq s} \| \, \partial^\alpha \theta \, \|_{L^p(X)}\,, \tag{2.47}$$

or with the equivalent norm

$$\| \, \theta \, \|_{W^{s,p}(X)} = \left(\sum_{|\alpha| \leq s} \| \, \partial^\alpha \theta \, \|^p_{L^p(X)} \right)^{\frac{1}{p}}\,. \tag{2.48}$$

The choice of which of the above norms to use depends upon which one allows for the easiest analysis of a given problem. The important point is that these two norms are equivalent, so that either may be used. For the case when $p = 2$, the space $W^{s,2}(X)$ is denoted by $H^s(X)$.

The definition of the Sobolev spaces as stated above requires that $s \geq 0$ be an integer. However, the definition can be extended in such a way that s need not be an integer, but that it only satisfy $s \geq 0$. A Sobolev space $W^{s,p}(X)$ in which s is not an integer is called a fractional Sobolev space. We will occasionally refer to these spaces, but for brevity their precise definition will be omitted.

The Sobolev spaces in which $p < \infty$, have the desirable property that their member functions can approximated arbitrarily well by functions from the space $C^\infty(X)$. The following density theorem is a precise statement of this fact.

Theorem 3. *Let $\theta \in W^{s,p}(X)$ where $s \geq 0$ and $1 \leq p < +\infty$ is an integer. Then, $\exists$ a sequence $\{\, \theta_n \,\}_{n \geq 0}$ in $C^\infty(X) \cap W^{s,p}(X)$ such that*

$$\theta_n \to \theta \ \ \textit{in}\ W^{s,p}(X)\,. \tag{2.49}$$

The following theorem is similar to the previous theorem, except that the compactly supported space $C_c^\infty(\mathbb{R}^d)$ is used instead $C^\infty(X)$.

Theorem 4. *Let $\theta \in W^{1,p}(X)$ with $1 \leq p < +\infty$. Then, $\exists$ a sequence $\{\theta_n\}_{n\geq 0}$ in $C_c^\infty(\mathbb{R}^d)$ such that*

$$(i)\ \ \theta_n \to \theta \ \ in\ L^p(X) \quad and \tag{2.50}$$

$$(ii)\ \ (\nabla\theta_n)_{|V} \to (\nabla\theta)_{|V} \ \ in\ [L^p(X)]^d,\ \forall\ V\ such\ that\ \overline{V} \subset X\ and\ \overline{V}\ is\ compact. \tag{2.51}$$

We remark that in this theorem the approximating sequence can only be said to be in the compactly supported space $C_c^\infty(\mathbb{R}^d)$, and not in the space $C_c^\infty(X)$.

In this work, we want to consider broken Sobolev functions in which it is meaningful to speak of their local boundary values. To find conditions that will guarantee such functions have local boundary values requires that we first understand when it is meaningful to speak of the boundary values of the usual Sobolev functions. To this end, let $\gamma_0 : C^0(\overline{X}) \to C^0(\partial X)$ be the standard trace operator for continuous functions. This operator can be extended in a unique way to the Sobolev space $W^{1,p}(\Omega)$, for $s \geq 1$ and $p < \infty$, provided that the domain X is a Lipschitz bounded open set. This result is stated in the following theorem. A proof of this theorem is given in [2].

Theorem 5. *Let $1 \leq p < +\infty$ and let X be a Lipschitz bounded open set. Then the trace operator $\gamma_0 : C^0(\overline{X}) \to C^0(\partial X)$ can be uniquely extended to $W^{1,p}(X)$. Moreover, it satisfies*

$$(i)\ \gamma_0 : W^{1,p}(X) \to W^{\frac{1}{p'},p}(\partial X)\ \ is\ surjective\ and \tag{2.52}$$

$$(ii)\ the\ nullspace\ of\ \gamma_0\ is\ W_0^{1,p}(X)\,, \tag{2.53}$$

where $W_0^{1,p}(X)$ is the closure of $C_c^\infty(X)$ with respect to $W^{1,p}(X)$ and p' satisfies $\frac{1}{p} + \frac{1}{p'} = 1$.

In the case $p = 2$, we note that $\gamma_0 : H^1(X) \to H^{\frac{1}{2}}(\partial X)$. In general, if $s \geq 1/2$, then the above trace theorem can be used to show that γ_0 is surjective mapping from $H^s(\Omega)$ to $H^{s-\frac{1}{2}}(\partial\Omega)$, and more generally, γ_0 is surjective mapping from $W^{s,p}(\Omega)$ to $W^{s-\frac{1}{p'},p}(\partial\Omega)$, for $s \geq 1/p'$. For a given function θ, we hereafter denote its trace $\gamma_0(\theta)$ by $\theta_{|\partial X}$.

A useful inverse inequality for functions $\theta \in W_0^{1,p}(\Omega)$ is the following:

Lemma 3. [Poincare′s Inequality] *Let $1 \leq p < +\infty$ and let $\Omega \subset \mathbb{R}^d$ be a bounded, open*

set. Then $\exists\ C = C(\,p, \Omega\,) > 0$ *such that*

$$\| \theta \|_{L^p(\Omega)} \ \leq \ C \, \| \nabla \theta \|_{L^p(\Omega)} \tag{2.54}$$

holds, $\forall\ \theta \in W_0^{1,p}(\Omega)$.

The last Sobolev space result to be presented is the following set of inverse inequalities [8] for functions $\theta \in H^s(\Omega)$:

Lemma 4. *Let* $\Omega \subset \mathbb{R}^3$. *Then*

$$(i) \ \ \forall\, K_j \in \mathcal{T}_h,\ \forall\ \theta \in H^1(K_j), \qquad \| \theta \|^2_{0,f_k} \leq C \left(\frac{1}{h_j} | \theta |^2_{0,K_j} + h_j | \theta |^2_{0,K_j} \right), \qquad \text{and} \tag{2.55}$$

$$(ii) \ \ \forall\, K_j \in \mathcal{T}_h,\ \forall\ \theta \in H^2(K_j), \qquad \| \nabla\theta \|^2_{0,f_k} \leq C \left(\frac{1}{h_j} | \theta |^2_{1,K_j} + h_j | \theta |^2_{2,K_j} \right), \tag{2.56}$$

where for each K_j, $f_k \in \mathcal{F}_h$ *is such that* $f_k \subset \partial K_j$.

Broken Sobolev spaces

We now define the broken Sobolev spaces. It is important to note that by their very definition these spaces are mesh dependent.

Definition 5. *Let* $1 \leq p \leq +\infty$ *and* $s \geq 0$ *be integers. The broken Sobolev space* $W^{s,p}(\mathcal{T}_h)$ *is defined as*

$$W^{s,p}(\mathcal{T}_h) \ := \ \{\, \theta \in L^p(\Omega) : \ \theta_{|K_j} \in W^{s,p}(K_j),\ j = 1, \ldots, N_h \,\}, \tag{2.57}$$

Moreover, this space is a Banach space when equipped with the norm

$$\| \theta \|_{W^{s,p}(\mathcal{T}_h)} = \left(\sum_{K_j \in \mathcal{T}_h} \| \theta \|^p_{W^{s,p}(K_j)} \right)^{\frac{1}{p}}. \tag{2.58}$$

From this definition, it is seen that broken Sobolev functions are locally Sobolev functions, i.e., if $\theta \in W^{s,p}(\mathcal{T}_h)$, then $\theta_{|K_j} \in W^{s,p}(K_j),\ \forall\, K_j \in \mathcal{T}_h$. The space $W^{s,p}(\mathcal{T}_h)$ is denoted by $H^s(\mathcal{T}_h)$, when $p = 2$.

The reason for referring to functions in $W^{s,p}(\mathcal{T}_h)$ as "broken" functions can be understood by considering the implications of Theorem 5 to this space of functions. In particular, Theorem 5 implies that for the space $W^{s,p}(\mathcal{T}_h),\ s \geq 1/p'$, there exists a family of local trace

operators $\{\gamma_{0,j}\}_{j=1}^{N_h}$, where $\gamma_{0,j} : W^{s,p}(K_j) \rightarrow W^{\frac{1}{p'},p}(\partial K_j)$ is surjective, for $k = 1, \ldots, N_h$. Now, let f_k be an interior face, where $f_k = K_1 \cap K_2$. If $\theta \in W^{s,p}(\mathcal{T}_h)$, then we denote the value of $\gamma_{0,1}(\theta)$ on f_k by $(\theta_{|K_1})_{|f_k}$ and the value of $\gamma_{0,2}(\theta)$ on f_k by $(\theta_{|K_2})_{|f_k}$. Thus, we see it is not meaningful to speak of the value of θ on f_k, since θ is multi-valued on f_k. It has the value $(\theta_{|K_1})_{|f_k}$ if you approach f_k from the interior of K_1 and it has the value $(\theta_{|K_2})_{|f_k}$ if you approach f_k from the interior of K_2. It only makes sense to speak of the value of θ on the face f_k when $f_k \subset \partial\Omega$.

We now introduce two families of operators that are defined on $W^{s,p}(\mathcal{T}_h),\ s \geq 1/p'$. The first family of operators gives the average value of functions in $W^{s,p}(\mathcal{T}_h)$ across the interior faces. The second family of operators gives the jump in the values of a functions in $W^{s,p}(\mathcal{T}_h)$ across the interior faces.

Definition 6. *Let $1 \leq p < +\infty$ be an integer and let $s \geq 1/p'$. For $k = 1, \ldots, P_h$, we define the operators $\{\cdot\}_k : W^{s,p}(\mathcal{T}_h) \rightarrow \mathbb{R}$ and $[\cdot]_k : W^{s,p}(\mathcal{T}_h) \rightarrow \mathbb{R}$ as follows: let K_1 and K_2 be the distinct elements in $\mathcal{T}_h$ such that $f_k = K_1 \cap K_2$, where $\nu_{K_1} = \nu_k$ on f_k and $\nu_{K_2} = -\nu_k$ on f_k. Then*

$$\{\,\theta\,\}_k \;=\; \frac{1}{2}\left(\,\theta_{|K_1} \;+\; \theta_{|K_2}\,\right), \qquad \forall\ \theta \in W^{s,p}(\mathcal{T}_h), \tag{2.59}$$

and

$$[\,\theta\,]_k \;=\; \theta_{|K_1} \;-\; \theta_{|K_2}\,, \qquad \forall\ \theta \in W^{s,p}(\mathcal{T}_h). \tag{2.60}$$

where by $\theta_{|K_1}$ and $\theta_{|K_2}$ we mean $(\theta_{|K_1})_{|f_k}$ and $(\theta_{|K_2})_{|f_k}$.

The family $\{\,\{\cdot\}_k\,\}_{k=1}^{P_h}$ is known as the family of average operators and the family $\{\,[\cdot]_k\,\}_{k=1}^{P_h}$ is known as the family of jump operators on the interior faces. We will find it convenient to write the jump operator $\{\cdot\}_k$ and the average operator $[\cdot]_k$ as $\{\cdot\}$ and $[\cdot]$, respectively, since it will always be clear from the context what interior face f_k is being considered.

The average and jump operators may be used to greatly simplify notation during the derivation of DG formulations. The ability to use these operators is a result of the following well-known identity.

Lemma 5. *Let $1 \leq p < +\infty$ be an integer and let $s \geq 1/p'$. Let $\xi \in [W^{s,p}(\mathcal{T}_h)]^3$ and let*

$\theta \in W^{s,p}(\mathcal{T}_h)$. *Then*

$$\sum_{j=1}^{N_h} \int_{\partial K_j} \xi_{|K_j} \cdot \nu_{K_j}\, \theta_{|K_j}\, dS \;=\; \sum_{f_k \in \partial\Omega} \int_{f_k} \xi_{|K_j} \cdot \nu_k\, \theta_{|K_j}\, dS + \sum_{k=1}^{P_h} \int_{f_k} (\,\{\xi \cdot \nu_k\}[\,\theta\,] \;+\; [\xi \cdot \nu_k]\{\,\theta\,\}\,)\; dS\,. \tag{2.61}$$

Proof. For any given interior face f_k, we assume that $f_k = K_1 \cap K_2$ and that the outward unit normal to K_1, ν_{K_1}, is equal to ν_k. Then

$$\begin{aligned}
&\sum_{j=1}^{N_h} \int_{\partial K_j} \xi_{|K_j} \cdot \nu_{K_j}\, \theta_{|K_j}\, dS \\
&\quad= \sum_{\nu_k \in \partial\Omega} \int_{f_k} \xi \cdot \nu_k\, \theta\, dS \;+\; \sum_{k=1}^{P_h} \int_{f_k} (\,\xi_{|K_1} \cdot \nu_{K_1}\, \theta_{|K_1} + \xi_{|K_2} \cdot \nu_{K_2}\, \theta_{|K_2}\,)\, dS \\
&\quad= \sum_{\nu_k \in \partial\Omega} \int_{f_k} \xi \cdot \nu_k\, \theta\, dS \;+\; \sum_{k=1}^{P_h} \int_{f_k} (\,\xi_{|K_1} \cdot \nu_k\, \theta_{|K_1} - \xi_{|K_2} \cdot \nu_k\, \theta_{|K_2}\,)\, dS \\
&\quad= \sum_{\nu_k \in \partial\Omega} \int_{f_k} \xi \cdot \nu_k\, \theta\, dS \;+\; \sum_{k=1}^{P_h} \int_{f_k} [\,\xi \cdot \nu_k\, \theta\,]\, dS\,.
\end{aligned} \tag{2.62}$$

Now let a_1, b_1, a_2, $b_2 \in \mathbb{R}$. Then

$$\begin{aligned}
&\frac{1}{2}\,(a_1 + a_2)\,(b_1 - b_2) \;+\; (a_1 - a_2)\,\frac{1}{2}\,(b_1 + b_2) \\
&\qquad= \;\frac{1}{2}\,(\,a_1\,b_1 - a_1\,b_2 + a_2\,b_1 - a_2\,b_2\,) \;+\; \frac{1}{2}\,(\,a_1\,b_1 + a_1\,b_2 - a_2\,b_1 - a_2\,b_2\,) \\
&\qquad= \;a_1\,b_1 \;-\; a_2\,b_2\,.
\end{aligned} \tag{2.63}$$

Upon setting $a_1 = \xi_{|K_1} \cdot \nu_k$, $a_2 = \xi_{|K_2} \cdot \nu_k$, $b_1 = \theta_{|K_1}$, and $b_2 = \theta_{|K_2}$ in (2.63), we get that

$$\{\,\xi \cdot \nu_k\,\}\,[\,\theta\,] \;+\; [\,\xi \cdot \nu_k\,]\,\{\,\theta\,\} \;=\; [\,\xi \cdot \nu_k\,\theta\,]\,. \tag{2.64}$$

Using this identity in (2.62) leads to (2.61). ∎

2.2.3 Broken approximation space $D_r(\mathcal{T}_h)$

With the $W^{s,p}(\mathcal{T}_h)$ well defined, we now turn our attention to defining a discrete, or finite-dimensional, space $D_r(\mathcal{T}_h)$, whose functions will be used to approximate functions in $W^{s,p}(\mathcal{T}_h)$. The discrete space will be chosen so that it satisfies $D_r(\mathcal{T}_h) \subset W^{s,p}(\mathcal{T}_h)$ and so that it is locally a polynomial space. Thus, in order to define $D_r(\mathcal{T}_h)$, we must first define the polynomial spaces.

Polynomial spaces

Let $r \geq 0$ be an integer. For an arbitrary domain $X \subset \mathbb{R}^d,\ d \in \mathbb{N}$, the standard space of polynomials $\mathbb{P}^r(X)$ on X is defined as

$$\mathbb{P}^r(X) \ = \ \{\ p \in L^2(X) : p(x) = \sum_{\substack{0 \leq i_1,\ldots,i_d \leq r \\ i_1+\ldots+i_d \leq r}} \alpha_{i_1,\ldots,i_d}\, x_1^{i_1} \cdots x_d^{i_d}\ ,\ \alpha_{i_1,\ldots,i_d} \in \mathbb{R}\ \} \tag{2.65}$$

and the space of tensor product polynomials $\mathbb{Q}^r(X)$ on X is defined by

$$\mathbb{Q}^r(X) \ = \ \{\ p \in L^2(X) : p(x) = \sum_{0 \leq i_1,\ldots,i_d \leq r} \alpha_{i_1,\ldots,i_d}\, x_1^{i_1} \cdots x_d^{i_d}\ ,\ \alpha_{i_1,\ldots,i_d} \in \mathbb{R}\ \}\,. \tag{2.66}$$

It clearly follows from these definitions that $\mathbb{P}^r(X) \subset \mathbb{Q}^r(X)$. For $d = 3$, it is easy to show that $\dim \mathbb{P}^r(X) = \frac{1}{6}(r+1)(r+2)(r+3)$ and $\dim \mathbb{Q}^r(X) = (r+1)^3$.

One common choice of basis for $\mathbb{Q}^r([\,0,1\,])$ is the well-known Legendre polynomial basis $\{\, L_0(x), \ldots, L_r(x)\,\}$. These functions are pairwise orthogonal and satisfy

$$\begin{aligned}
&(i) \quad \| L_m \|^2_{0,(0,1)} \ = \ \frac{1}{2m+1}\ , \\
&(ii) \quad L_m(1) \ = \ 1\ , \qquad\qquad \text{and} \\
&(iii) \quad L_m(0) \ = \ (-1)^m\ , \qquad \forall\ m = 0, \ldots, r\,.
\end{aligned}$$

The Legendre basis for $\mathbb{Q}^r(\widehat{K})$ is defined to be $\{\, L_i\, L_j\, L_k\,\}_{i,j,k=0}^{r}$.

Since $\mathbb{Q}^r(\widehat{K}) \subset H^2(\widehat{K})$, the inverse inequalities (2.55) are valid for functions in $\mathbb{Q}^r(\widehat{K})$. However, stronger inverse inequalities may be found by taking advantage of well-established properties of polynomials.

The following lemma is a new result. It is an improvement over a standard result that can

be found in [68]. The improvement is that the bounds it establishes scale with the optimal factor $r^{1/2}$, as opposed to the previously established scale factor of r.

Lemma 6. *Let $\hat{f}$ be an arbitrary face of $\widehat{K} = [0,1]^3$ and let $\hat{e}$ be an edge of $\widehat{K}$ such that $\hat{e} \subset \partial\hat{f}$. If $p \in \mathbb{Q}_r(\widehat{K})$, $r \geq 1$, then there exists a constant C, independent of r, such that*

$$\begin{aligned} \| p \|_{0,\hat{e}} &\leq C\, r^{1/2} \| p \|_{0,\hat{f}}\,, && (2.67)\\ \| p \|_{0,\hat{f}} &\leq C\, r^{1/2} \| p \|_{0,\widehat{K}}\,, && (2.68)\\ \| \nabla p \|_{0,\hat{e}} &\leq C\, r^{1/2} \| \nabla p \|_{0,\hat{f}}\,, && (2.69)\\ \| \nabla p \|_{0,\hat{f}} &\leq C\, r^{1/2} \| \nabla p \|_{0,\widehat{K}}\,. && (2.70) \end{aligned}$$

Moreover, these results are optimal in r.

Proof. For definiteness, let us assume that $\hat{e} = [0,1]\times\{1\}\times\{1\}$ and $\hat{f} = [0,1]\times[0,1]\times\{1\}$. To show that the above results are optimal, consider the polynomial $p(\hat{x}) = \hat{x}_1^r\,\hat{x}_2^r\,\hat{x}_3^r$. Then a few calculations shows that

$$\| p \|^2_{0,\widehat{K}} = \frac{1}{(2r+1)^3}\,, \qquad \| p \|^2_{0,\hat{f}} = \frac{1}{(2r+1)^2}\,, \qquad \| p \|^2_{0,\hat{e}} = \frac{1}{(2r+1)}\,.$$

Therefore,

$$\begin{aligned} \| p \|_{0,\hat{f}} &= \sqrt{2r+1}\, \| p \|_{0,\widehat{K}} \geq \sqrt{2r}\, \| p \|_{0,\widehat{K}} \qquad \text{and}\\ \| p \|_{0,\hat{e}} &= \sqrt{2r+1}\, \| p \|_{0,\hat{f}}\,. \geq \sqrt{2r}\, \| p \|_{0,\hat{f}}\,. \end{aligned} \tag{2.71}$$

That (2.69) and (2.70) are optimal follows from the above argument, since ∇p is a summation of polynomials in $\mathbb{Q}_r(\widehat{K})$, so that the above argument given for p applies to $\nabla\, p$ as well.

Now let $p(\,\hat{x}_1, \hat{x}_2, \hat{x}_3\,) = \sum_{i,j,k=0}^{r} a_{i,j,k} L_i(\hat{x}_1) L_j(\hat{x}_2) L_k(\hat{x}_3) \in \mathbb{Q}^r(\widehat{K})$ be arbitrary. Then it

follows that

$$
\begin{aligned}
&\| p \|^2_{0,\widehat{K}} \\
&\quad = \sum_{i,j,k=0}^{r} a_{i,j,k} \sum_{s,t,u=0}^{r} a_{s,t,u} \int\limits_{\widehat{K}} \big(L_i(\hat{x}_1) L_j(\hat{x}_2) L_k(\hat{x}_3) \big) \, \big(L_s(\hat{x}_1) L_t(\hat{x}_2) L_u(\hat{x}_3) \big) \, d\hat{x}_1 d\hat{x}_2 d\hat{x}_3 \\
&\quad = \sum_{i,j,k=0}^{r} a_{i,j,k} \sum_{s,t,u=0}^{r} a_{s,t,u} \int\limits_{(0,1)} L_i L_s \, d\hat{x}_1 \int\limits_{(0,1)} L_j L_t \, d\hat{x}_2 \int\limits_{(0,1)} L_k L_u \, d\hat{x}_3 \\
&\quad = \sum_{i,j,k=0}^{r} \frac{a^2_{i,j,k}}{(2i+1)(2j+1)(2k+1)} \, . \qquad (2.72)
\end{aligned}
$$

To bound $\| p \|_{0,\hat{f}}$, we use the fact that $p_{|\hat{f}} = \sum\limits_{i,j,k=0}^{r} a_{i,j,k} L_i(\hat{x}_1) L_j(\hat{x}_2)$. Thus,

$$
\begin{aligned}
\| p \|^2_{0,\hat{f}} &= \sum_{i,j,k=0}^{r} \sum_{s,t,u=0}^{r} \int\limits_{(0,1)^2} \big(a_{i,j,k} \, L_i(\hat{x}_1) L_j(\hat{x}_2) \big) \, \big(a_{s,t,u} \, L_s(\hat{x}_1) L_t(\hat{x}_2) \big) \, d\hat{x}_1 d\hat{x}_2 \\
&= \sum_{i,j,k=0}^{r} \sum_{u=0}^{r} \frac{a_{i,j,k} \, a_{i,j,u}}{(2i+1)(2j+1)} \\
&= \sum_{i,j=0}^{r} \frac{1}{(2i+1)(2j+1)} \sum_{k,u=0}^{r} a_{i,j,k} \, a_{i,j,u} \\
&\le \sum_{i,j=0}^{r} \frac{1}{(2i+1)(2j+1)} \left(\sum_{k=0}^{r} a^2_{i,j,k} \right)^{1/2} \left(\sum_{u=0}^{r} a^2_{i,j,u} \right)^{1/2} \\
&= \sum_{i,j=0}^{r} \frac{1}{(2i+1)(2j+1)} \sum_{k=0}^{r} a^2_{i,j,k} \\
&\le \sum_{i,j=0}^{r} \frac{1}{(2i+1)(2j+1)} \sum_{k=0}^{r} a^2_{i,j,k} \left(\frac{2r+1}{2k+1} \right) \\
&\le (2r+1) \sum_{i,j,k=0}^{r} \frac{a^2_{i,j,k}}{(2i+1)(2j+1)(2k+1)} \\
&= (2r+1) \, \| p \|^2_{0,\widehat{K}} \\
&\le 3r \, \| p \|^2_{0,\widehat{K}} \, . \qquad (2.73)
\end{aligned}
$$

In order to bound $\| p \|_{0,\hat{e}}$, we first write $p_{|\hat{f}} = \sum\limits_{i,j=0}^{r} b_{i,j} L_i(\hat{x}_1) L_j(\hat{x}_2)$, where $b_{i,j} = \sum_{k=0}^{r} a_{i,j,k}$.

It then easily follows that

$$\begin{aligned} \| p \|^2_{0,\hat{f}} &= \sum_{i,j=0}^{r} b_{i,j} \sum_{s,t=0}^{r} b_{s,t} \int\limits_{(0,1)} L_i(\hat{x}_1) L_s(\hat{x}_2)\, d\hat{x}_1 \int\limits_{(0,1)} L_j(\hat{x}_1) L_t(\hat{x}_2)\, d\hat{x}_2 \\ &= \sum_{i,j=0}^{r} \frac{b_{i,j}^2}{(2i+1)(2j+1)} \,. \end{aligned} \tag{2.74}$$

Using the fact that $p_{|\hat{e}} = \sum_{i,j=0}^{r} b_{i,j} L_i(x_1)$, we get that

$$\begin{aligned} \| p \|^2_{0,\hat{e}} &= \sum_{i,j}^{r} b_{i,j} \sum_{s,t=0}^{r} b_{s,t} \int\limits_{(0,1)} L_i(\hat{x}_1)\; L_s(\hat{x}_1)\, d\hat{x}_1 \\ &= \sum_{i,j=0}^{r} \sum_{t=0}^{r} \frac{b_{i,j}\, b_{i,t}}{2i+1} \\ &= \sum_{i=0}^{r} \frac{1}{2i+1} \sum_{j,t=0}^{r} b_{i,j}\, b_{i,t} \\ &\leq \sum_{i=0}^{r} \frac{1}{2i+1} \left(\sum_{j=0}^{r} b_{i,j}^2 \right)^{1/2} \left(\sum_{t=0}^{r} b_{i,t}^2 \right)^{1/2} \\ &= \sum_{i=0}^{r} \frac{1}{2i+1} \sum_{j=0}^{r} b_{i,j}^2 \\ &= \sum_{i=0}^{r} \frac{1}{2i+1} \sum_{j=0}^{r} b_{i,j}^2 \left(\frac{2r+1}{2j+1} \right) \\ &= (2r+1) \sum_{i,j=0}^{r} \frac{b_{i,j}^2}{(2i+1)(2j+1)} \\ &= (2r+1) \| p \|^2_{0,\hat{f}} \\ &\leq 3r \| p \|^2_{0,\hat{f}} \,. \end{aligned} \tag{2.75}$$

To prove (2.69) and (2.70), we can mimic the same argument used to prove (2.67) and (2.68). To do this, we first note that since $(\partial/\partial\, \hat{x}_d)\; p \in \mathbb{Q}^r(\widehat{K})$, for $d = 1, 2, 3$, it follows that there exists coefficients $\beta^d_{i,j,k}$ such that

$$\frac{\partial}{\partial\, \hat{x}_d} p \;=\; \sum_{i,j,k=0}^{r} \beta^d_{i,j,k} L_i(\hat{x}_1) L_j(\hat{x}_2) L_k(\hat{x}_3) \,. \tag{2.76}$$

By applying the previous argument to each of the functions $(\partial/\partial\,\hat{x}_d)\,p$, we end up with $\|\,\nabla p\,\|^2_{0,\hat{f}} \;\leq\; 9r\,\|\,\nabla p\,\|^2_{0,\hat{K}}$ and $\|\,\nabla p\,\|^2_{0,\hat{e}} \;\leq\; 9r\,\|\,\nabla p\,\|^2_{0,\hat{f}}$. □

Broken polynomial spaces

For a given integer $r \geq 0$, the discrete approximation space $D_r(\mathcal{T}_h)$ is defined to be

$$D_r(\mathcal{T}_h) \;=\; \{\, w_h \in L^2(\Omega) :\, (\,w_h\,)_{|K_j} \in \mathbb{Q}^r(K_j),\; j = 1,\ldots,N_h \,\}\,. \tag{2.77}$$

From this definition, it easy to see that the broken approximation functions in $D_r(\mathcal{T}_h)$ are locally polynomial functions. Since $D_r(\mathcal{T}_h) \subset W^{s,p}(\mathcal{T}_h)$, for any integers $s \geq 0$ and $1 \leq p \leq \infty$, any properties that hold for $W^{s,p}(\mathcal{T}_h)$ also hold for $D_r(\mathcal{T}_h)$.

For each element K_j, a local basis $\{\psi_0,\ldots,\psi_r\}$ for $\mathbb{Q}^r(K_j)$ can be generated via the reference basis $\{\hat{\psi}_0,\ldots,\hat{\psi}_r\}$ by defining $\psi_m(x) = \hat{\psi}_m(\,T^{-1}_{K_j}(x)\,)$, for $m = 0,\ldots,r$, where $T^{-1}_{K_j}(x) = M^{-1}_{K_j}x - M^{-1}_{K_j}b_{K_j}$. Thus, there is a one-to-one correspondence between functions in $\mathbb{Q}^r(K_j)$ and $\mathbb{Q}^r(\widehat{K})$.

We now take advantage of the one-to-one correspondence between $\mathbb{Q}^r(K_j)$ and $\mathbb{Q}^r(\widehat{K})$ to derive some important relationships that hold for functions from these spaces. Let e_{K_j} and f_{K_j} be an arbitrary edge and face of some element K_j. If $p \in \mathbb{Q}^r(K_j)$ and $\hat{p}$ is the image of p in the reference element, i.e., $\hat{p} = p \circ T^{-1}_{K_j}$, then

$$\begin{aligned}\|\,p\,\|_{0,K_j} \;&=\; \left(\int_{K_j} p(x)\,p(x)\,dx\right)^{1/2} \;=\; \left(\int_{\widehat{K}} \hat{p}(\hat{x})\,\hat{p}(\hat{x})\,|\det(M_{K_j})|\,d\hat{x}\right)^{1/2}\\ &=\; |\,K_j\,|^{1/2}\,\|\,\hat{p}\,\|_{0,\widehat{K}}\,,\end{aligned} \tag{2.78}$$

Using a similar argument, it can also be shown that $\|\,p\,\|_{0,f_{K_j}} \;=\; |\,f_{K_j}\,|^{1/2}\,\|\,\hat{p}\,\|_{0,\hat{f}}$ and $\|\,p\,\|_{0,e_{K_j}} = |\,e_{K_j}\,|^{1/2}\,\|\,\hat{p}\,\|_{0,\hat{e}}$.

Taking into account Lemma 2, along with the above equalities, yields that

$$\|\,p\,\|_{0,K_j} \;\sim\; h_j^{3/2}\,\|\,\hat{p}\,\|_{0,\widehat{K}}\,, \tag{2.79}$$

$$\|\,p\,\|_{0,f_{K_j}} \;\sim\; h_j\,\|\,\hat{p}\,\|_{0,\hat{f}}\,, \tag{2.80}$$

$$\|\,p\,\|_{0,e_{K_j}} \;\sim\; h_j^{1/2}\,\|\,\hat{p}\,\|_{0,\hat{e}}\,. \tag{2.81}$$

To find a relationship between the gradients of p and $\hat{p}$, we first note that $\nabla p = (M_{K_j}^{-1})^T \nabla \hat{p}$ and $\nabla \hat{p} = M_{K_j}^T \nabla p$. These inequalities imlpy that $|\nabla p| \leq \| M_{K_j}^{-1} \| \, |\nabla \hat{p}|$ and $\frac{1}{\| M_{K_j} \|} |\nabla \hat{p}| \leq |\nabla p|$. From Lemma 1, we have that $\| M_{K_j}^{-1} \| \leq 3\,\rho_j^{-1}$, which implies $\| M_{K_j}^{-1} \| \leq 3\,h_j^{-1}$, and that $\frac{6}{\pi}\, h_j^{-1} \leq \frac{1}{\| M_{K_j} \|}$. Thus,

$$\frac{\pi}{6}\, h_j^{-1}\, |\nabla \hat{p}| \;\leq\; |\nabla p| \;\leq\; 3 h_j^{-1}\, |\nabla \hat{p}|\,, \tag{2.82}$$

which shows that $|\nabla p| \sim h_j^{-1}\, |\nabla \hat{p}|$.

The analogous inequalities to (2.79)-(2.81) for ∇p and ∇p are now easily shown to be

$$\| \nabla p \|_{0,K_j} \;\sim\; h_j^{1/2}\, \| \nabla \hat{p} \|_{0,\widehat{K}}\,, \tag{2.83}$$

$$\| \nabla p \|_{0,f_{K_j}} \;\sim\; \| \nabla \hat{p} \|_{0,\hat{f}}\,, \tag{2.84}$$

$$\| \nabla p \|_{0,e_{K_j}} \;\sim\; h_j^{-1/2}\, \| \nabla \hat{p} \|_{0,\hat{e}}\,. \tag{2.85}$$

Inverse inequalities for $\mathbb{Q}^r(K_j)$ are now established by combining equivalencies (2.79)-(2.81) and (2.83)-(2.85) with Lemma 6. The following well known lemma states these results.

Lemma 7. *Let $K_j \subset \mathbb{R}^3$ be an arbitrary element in $\mathcal{T}_h$. Let f_{K_j} be an arbitrary face of K_j and let e_{K_j} be an edge of K_j such that $e_{K_j} \subset \partial f_{K_j}$. If p is a polynomial of degree $r \geq 1$ on K_j, i.e., $p \in \mathbb{Q}_r(K_j)$, then there exists a constant C, independent of h_j and r, such that*

$$\| p \|_{0,e_{K_j}} \;\leq\; C\, h_j^{-1/2}\, r^{1/2}\, \| p \|_{0,f_{K_j}}\,, \tag{2.86}$$

$$\| p \|_{0,f_{K_j}} \;\leq\; C\, h_j^{-1/2}\, r^{1/2}\, \| p \|_{0,K_j}\,, \tag{2.87}$$

$$\| \nabla p \|_{0,e_{K_j}} \;\leq\; C\, h_j^{-1/2}\, r^{1/2}\, \| \nabla p \|_{0,f_{K_j}}\,, \tag{2.88}$$

$$\| \nabla p \|_{0,f_{K_j}} \;\leq\; C\, h_j^{-1/2}\, r^{1/2}\, \| \nabla p \|_{0,K_j}\,. \tag{2.89}$$

Proof. Let K_j be any arbitrary element in $\mathcal{T}_h$ and let e_{K_j} be any edge of K_j and let f_{K_j} be any face of K_j such that $e_{K_j} \subset \partial f_{K_j}$. Let $\hat{e}$ and $\hat{f}$ be the images of e_{K_j} and f_{K_j} in the reference element $\widehat{K}$. For a given $p \in \mathbb{Q}^r(K_j)$, let $\hat{p}$ be the image of p in $\widehat{K}$.

Using (2.81), (2.67) from Lemma 6, and then (2.80) leads to the inequality

$$\| p \|_{0,e_{K_j}} \;\leq\; C\, h_j^{1/2}\, \| \hat{p} \|_{0,\hat{e}} \;\leq\; C\, h_j^{1/2} r^{1/2}\, \| \hat{p} \|_{0,\hat{f}} \;\leq\; C\, h_j^{-1/2} r^{1/2}\, \| p \|_{0,f_{K_j}}\,. \tag{2.90}$$

Similarly, using (2.80), (2.68) from Lemma 6, and (2.79) gives that

$$\| p \|_{0,f_{K_j}} \;\leq\; C\, h_j \| \hat{p} \|_{0,\hat{f}} \;\leq\; C\, h_j r^{1/2} \| \hat{p} \|_{0,\widehat{K}} \;\leq\; C\, h_j^{-1/2} r^{1/2} \| p \|_{0,K_j} \,. \tag{2.91}$$

Likewise for ∇p, we can use (2.85), (2.69) from Lemma 6, and then (2.84) to end up with

$$\| \nabla p \|_{0,e_{K_j}} \;\leq\; C\, h_j^{-1/2} \| \nabla \hat{p} \|_{0,\hat{e}} \;\leq\; C\, h_j^{-1/2} r^{1/2} \| \nabla \hat{p} \|_{0,\hat{f}} \;\leq\; C\, h_j^{-1/2} r^{1/2} \| \nabla p \|_{0,f_{K_j}} \,. \tag{2.92}$$

Lastly, using (2.84), (2.70) from Lemma 6, and (2.83) lead

$$\| \nabla p \|_{0,f_{K_j}} \;\leq\; C \| \nabla \hat{p} \|_{0,\hat{f}} \;\leq\; C\, r^{1/2} \| \nabla \hat{p} \|_{0,\widehat{K}} \;\leq\; C\, h_j^{-1/2} r^{1/2} \| \nabla p \|_{0,K_j} \,. \tag{2.93}$$

□

2.2.4 Interpolation properties of $D_r(\mathcal{T}_h)$ to $H^s(\mathcal{T}_h)$.

The *hp*-interpolation properties of $D_r(\mathcal{T}_h)$ to $H^s(\mathcal{T}_h)$ are of fundamental importance, since these properties place a limit on how well any function from $D_r(\mathcal{T}_h)$ can approximate an unknown function from the broken Sobolev space $H^s(\mathcal{T}_h)$. The following theorem is a precise statement of these *hp*-interpolation properties and its proof was given in [12], [13].

Theorem 6. *Let $\Omega \subset \mathbb{R}^3$ be a polygonal domain. Let $\theta \in H^s(\mathcal{T}_h)$ and let $K_j \in \mathcal{T}_h$ be arbitrary. Let f_k be any arbitray face lying on ∂K_j. Then there exists a constant $C = C(s, \tau, \rho)$, independent of θ, r, and h, and an interpolation operator $\Pi_h^r : H^s(\mathcal{T}_h) \to D^r(\mathcal{T}_h)$ such that for any $0 \leq q \leq s$, the following inequalities hold:*

$$\| \theta - \Pi_h^r \theta \|_{q,K_j} \;\leq\; C\, \frac{h_j^{\mu-q}}{r^{s-q}} \| \theta \|_{s,K_j} \,, \qquad s \geq 0 \,, \tag{2.94}$$

$$\| \theta - \Pi_h^r \theta \|_{0,f_{K_j}} \;\leq\; C\, \frac{h_j^{\mu-\frac{1}{2}}}{r^{s-\frac{1}{2}}} \| \theta \|_{s,K_j} \,, \qquad s > 1/2 \,, \tag{2.95}$$

$$\| \theta - \Pi_h^r \theta \|_{1,f_{K_j}} \;\leq\; C\, \frac{h_j^{\mu-\frac{3}{2}}}{r^{s-\frac{3}{2}}} \| \theta \|_{s,K_j} \,, \qquad s > 3/2 \,, \tag{2.96}$$

where μ is defined to be

$$\mu \;=\; \min \{\, r+1,\, s \,\} \,.$$

Without loss of generality, Π_h^r can be taken to be the L^2-projection operator, since we are working in the space $H^s(\mathcal{T}_h)$. We also note that p is the standard notation used for the degree of the polynomials used to define $D^r(\mathcal{T}_h)$. However, in this work the variable r will be used instead of p.

Remark 1. *Theorem 6 is only valid under the assumptions placed on the domain Ω. In particular, for a domain that is a subset of $\mathbb{R}^4$ or higher, we are only guaranteed to have h-interpolation properties, which are given below.*

Remark 2. *Although Theorem 6 gives the existence of a function $\Pi_h^r\theta \in D^r(\mathcal{T}_h)$ satisfying properties (2.94) − (2.96) for a given function $\theta \in H^s(\mathcal{T}_h)$, $s \geq 0$, there does not exist a way of finding the interpolating function when θ is unknown. The problem considered throughout this work is that θ is unknown, but satisfies some given boundary-value partial differential equation. The goal is to develop a procedure using this differential equation that gives a constructive way of finding a unique function in $\theta_h \in D^r(\mathcal{T}_h)$ that approximates the unknown function θ. Moreover, it is desirable that the method developed is such that it can be used for finding error estimates for the difference $\theta - \theta_h$ that are similar to those in Theorem 6. In the best case scenario, the error estimates one finds will be optimal in both h and in r, which is to say that they attain the same order of convergence in h and r as do the estimates for the interpolanting function $\Pi_h^r\,\psi$.*

We now state a more general interpolation theorem for functions in $H^2(\mathcal{T}_h)$. This theorem relaxes the assumptions on the domain Ω. However, this does come at the cost of assuming that the polynomial degree r is a fixed, positive integer.

Theorem 7. *Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be a polygonal domain. Let $\theta \in H^s(\mathcal{T}_h)$ and let $K_j \in \mathcal{T}_h$ be arbitrary. Let f_k be any arbitray face lying on ∂K_j. Then there exists a constant $C = C(s, r, \tau, \rho)$, independent of θ and h, and an interpolation operator $\Pi_h^r : H^s(\mathcal{T}_h) \to D^r(\mathcal{T}_h)$ such that for any $0 \leq q \leq s$, the following inequalities hold:*

$$\| \theta - \Pi_h^r\theta \|_{q,K_j} \leq C\, h_j^{\mu-q} \| \theta \|_{s,K_j}\,, \qquad s \geq 0\,, \tag{2.97}$$

$$\| \theta - \Pi_h^r\theta \|_{0,f_{K_j}} \leq C\, h_j^{\mu-\frac{1}{2}} \| \theta \|_{s,K_j}\,, \qquad s > 1/2\,, \tag{2.98}$$

$$\| \theta - \Pi_h^r\theta \|_{1,f_{K_j}} \leq C\, h_j^{\mu-\frac{3}{2}} \| \theta \|_{s,K_j}\,, \qquad s > 3/2\,, \tag{2.99}$$

where μ is defined to be

$$\mu = \min\{\, r+1,\, s \,\}\,.$$

The next theorem we state is about the relationships that exist between between various Sobolev norms on the approximation space $D^r(\mathcal{T}_h)$. It is important to note that this theorem requires that the polynomial degree r be fixed, as is the case for Theorem 7, so that only $h-$refinement is permissible.

Theorem 8. *Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be a polygonal domain. Let $1 \leq q \leq +\infty$, $1 \leq p \leq +\infty$, and $0 \leq s \leq m$. Then $\exists\ C = C(r,q,p,s,m)$ such that $\forall\ \theta_h \in D^r(\mathcal{T}_h)$, we have that*

$$\| \theta_h \|_{W^{m,q}(K_j)} \ \leq \ C\, h^{s-m+d/q-d/p} \| \theta_h \|_{W^{s,p}(K_j)}\,, \quad \forall K_j \in \mathcal{T}_h\,. \tag{2.100}$$

In particular, for $m = 0$, $q = +\infty$, $s = 0$, and $p = 2$, the above estimate becomes

$$\| \theta_h \|_{L^\infty(K_j)} \ \leq \ C\, h^{-d/2} \| \theta_h \|_{0,K_j}\,, \quad \forall K_j \in \mathcal{T}_h\,. \tag{2.101}$$

The last theorem presented only holds since $D^r(\mathcal{T}_h)$ is a tensor product polynomial space. In this theorem, the constant C is independent of r.

Theorem 9. *Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be a polygonal domain. Let m and s be nonnegative integers such that $s - m - d/2 > 0$. Then $\exists\ C = C(s,m)$ such that, $\forall\ \theta \in H^s(\mathcal{T}_h)$, we have*

$$| \theta - \Pi_h^r \theta |_{W^{m,\infty}(K_j)} \ \leq \ C\, h^{s-m-d/2} \| \theta \|_{s,K_j}\,, \quad \forall K_j \in \mathcal{T}_h\,. \tag{2.102}$$

In particular, for $m = 0$, the above estimate becomes

$$\| \theta - \Pi_h^r \theta \|_{L^\infty(K_j)} \ \leq \ C\, h^{s-d/2} \| \theta \|_{s,K_j}\,, \quad \forall K_j \in \mathcal{T}_h\,. \tag{2.103}$$

We now derive one useful inequality, assuming that $s - d/2 > 0$, that results from combining the above two theorems.

$$\begin{aligned}
\|\theta - \theta_h\|_{L^\infty(K_j)} \ &\leq \ \|\theta - \Pi_h^r \theta\|_{L^\infty(K_j)} \ + \ \|\theta_h - \Pi_h^r \theta\|_{L^\infty(K_j)} \\
&\leq \ C h^{s-d/2} \| \theta \|_{s,K_j} \ + \ C\, h^{-d/2} \| \theta_h - \ \Pi_h^r \theta\|_{0,K_j} \\
&\leq \ C h^{s-d/2} \| \theta \|_{s,K_j} \ + \ C\, h^{-d/2} \| \theta - \ \Pi_h^r \theta\|_{0,K_j} \ + \ C\, h^{-d/2} \| \theta - \theta_h\|_{0,K_j} \\
&\leq \ C h^{s-d/2} \| \theta \|_{s,K_j} \ + \ C\, h^{\mu-d/2} \| \theta\|_{s,K_j} \ + \ C\, h^{-d/2} \| \theta - \theta_h\|_{0,K_j} \\
&\leq \ C\, h^{\mu-d/2} \| \theta\|_{s,K_j} \ + \ C\, h^{-d/2} \| \theta - \theta_h\|_{0,K_j}\,,
\end{aligned} \tag{2.104}$$

where the last line follows from the fact μ is defined to be $\mu \ = \ \min\{r+1, s\}$.

2.3 Useful Inequalities

This last section presents four inequalities that are of fundamental importance in the analysis of DG methods. These inequalities will be used repeatedly throughout this work, especially the first two that are given. The last two inequalities are crucial in proving time-dependent *a priori* error estimates for DG methods. The reader should note that explicit reference to the first two inequalities will not always be made when they are employed in the forthcoming analysis.

Given any constants $a,\, b,\, \epsilon > 0,$ the following two inequalities hold:

$$(i) \qquad (a + b)^2 \leq \frac{1}{2}a^2 + \frac{1}{2}b^2\,, \qquad (Young's\ inequality) \tag{2.105}$$

$$(ii) \qquad a\,b \leq \frac{1}{2\epsilon}a^2 + \frac{\epsilon}{2}b^2\,. \tag{2.106}$$

The next inequality is the well known Gronwall's lemma.

Lemma 8. *Suppose that the functions χ, B are in $C([\,0,T\,])$ and on nonnegative on $[0,T]$, If there A and C are nonnegative constants such that, $\forall\ 0 \leq t \leq T$,*

$$\chi(t) \leq A + C\int_0^t B(s)\,\chi(s)\,ds \tag{2.107}$$

is satisfied, then it follows that

$$\chi(t) \leq A\exp(\,C\int_0^t B(s)\,ds\,) \tag{2.108}$$

holds, $\forall\ t \in [\,0,T\,]$.

The following inequality [29] is similar in nature to Gronwall's lemma, except that it assumes that the integrand contains χ raised to the power 1 and the handside of the assumed inequality contains χ raised to the power 2. Thus, it yields a sharper bound than that secured by Gronwall's lemma.

Lemma 9. *Suppose that the functions χ, R, A, B are in $C([\,0,T\,])$, where $R, A,$ and B are nonnegative functions on $[\,0,T\,]$, as well. If C is a nonnegative constant such that, $\forall\ 0 \leq t \leq T$,*

$$\chi^2(t) + R(t) \leq A(t) + C\int_0^t B(s)\,\chi(s)\,ds \tag{2.109}$$

is satisfied, then it follows that

$$\sqrt{\chi^2(t) \; + \; R(t)} \; \leq \; \sup_{0 \leq s \leq t} A^{1/2}(s) \; + \; \frac{C}{2} \int_0^t B(s)\, ds \tag{2.110}$$

holds, $\forall \; t \in [\, 0, T \,]$.

Chapter 3

NIPG method of approximation to the potential

Discontinuous Galerkin methods are known to be very effective techniques for approximating the solutions of elliptic problems [40], [8], [15], [21], [23], [68], [64]. A comprehensive review of DG methods that have been developed for elliptic problems can be found in [9].

In particular, these methods should do well in numerically solving for the potential function arising in plasma systems, since this function satisfies Poisson's equation when the magnetic field is ignored. However, the fact that the plasma systems under consideration also satisfy the Vlasov equation must also be kept in mind, since the Vlasov equation is defined by the electric field $E(x,t)$, which is the gradient of the potential $\psi(x,t)$. Thus in the plasma setting, it is important that any DG method employed to approximate the Poisson equation also allows for approximation properties to be established for the electric field, not just the potential.

The nonsymmetric interior penalty method (NIPG) is a DG method for approximating the Poisson equation that was first introduced, and rigorously analyzed, in [68]. The NIPG method has the key properties that it yields an hp-error estimate for the electric field that is optimal in h, imposes both Dirichlet and Neumann boundary conditions weakly, and is such that it can be easily extended to new problems by incorporating additional penalty terms. For these reasons, this NIPG method is chosen in this work to approximate the Poisson equation arising in the context of a Vlasov-Poisson system.

The goals of this chapter are threefold. In section 1, the model Poisson problem will be introduced. A brief discussion of the existence, uniqueness, and regularity of solutions to

this problem will be discussed. In section 2, a comprehensive review of the NIPG method is given. This consists of a derivation of the weak formulation that defines the method, a discussion of the existence and uniqueness properties of both the infinite and finite dimensional NIPG problems, and a discussion of the *a priori hp*-error analysis results of the method. All of the results contained in section 2 were first given in [68]. In section 3, the original work on the NIPG method will be extended in three ways. First, the original *a priori* error estimate given in [68] will be improved so that it is optimal in both h and r, where r is the polynomial degree of the approximation, instead of just being optimal in h, Second, using the new *a priori* error estimate along with Lemma 6 and Theorem 6, *a priori* error estimates will be derived for new error quantities that will be needed in the DG analysis of the Vlasov-Poisson system. Third, an *hp*-error estimate will be proved for the error between the true solution to the Poisson system and the discrete solution given by the NIPG method to a perturbation of the Poisson system.

3.1 Poisson system

The model problem considered in this chapter is a second-order, elliptic boundary-value problem. The classical statement of the model problem is

$$
\begin{aligned}
-\nabla\cdot(A\nabla\psi) &= F \qquad \text{in } \Omega\,, && (3.1)\\
\psi &= r_{_D} \qquad \text{on } \Gamma_{_D}\,, && (3.2)\\
(A\nabla\psi)\cdot\nu &= r_{_N} \qquad \text{on } \Gamma_{_N}\,, && (3.3)
\end{aligned}
$$

where Ω is an open, bounded, polygonal domain $\subset \mathbb{R}^3$, $\Gamma_{_D}$ and $\Gamma_{_N}$ are disjoint subsets of $\partial\Omega$ and satisfy $\Omega = \Gamma_{_D} \cup \Gamma_{_N}$, ν is the outward, unit normal vector to $\partial\Omega$, and where $f \in L^2(\Omega)$, $r_{_D} \in L^2(\Gamma_{_D})$, and $r_{_N} \in L^2(\Gamma_{_N})$ are given functions. The condition $|\Gamma_{_D}| > 0$ will be assumed to hold throughout the remainder of this work. The assumptions on the matrix $A = (a_{i,j})_{i,j=1}^{3} : \Omega \to \mathbb{R}^{3\times 3}$ are that each $a_{i,j} \in C^1(\Omega)$, A is symmetric and semi-positive definite. Equations (3.1)-(3.3) taken together will be referred to as the Poisson system.

The existence and uniqueness properties of the Poisson equation (3.1)-(3.1) are studied in detail in [48], where the following result is established:

Theorem 10. *Under the above assumptions, there exists a unique solution, in a distributional sense, $\psi \in H^1(\Omega)$ satisfying the Poisson system (3.1)-(3.3).*

It is important to note that this theorem does not guarantee that the solution ψ to the

Poisson problem is locally an H^2-function. This fact plays a direct role in the consistency of the NIPG method for the Poisson problem, since in the formulation of the method it is assumed that ψ is locally an H^2-function.

3.2 NIPG method

We begin this section by deriving the NIPG formulation. Using this formulation, both the true and discrete NIPG variational problem statements are given. Then for completeness, a review of the current error estimates for the NIPG method is given. We will show exactly why these original estimates are sub-optimal in r.

Next, we will investigate in detail the sub-optimality in r of the NIPG method. Our investigation, combined with the improved inverse inequalities for polynomials given in Lemma 7, will lead to a proof that unequivocally shows that the error estimates for the NIPG method are in fact optimal in r, and hence are *hp*-optimal. To our knowledge, that the NIPG method is an *hp*-optimal method is a new result.

3.2.1 Weak formulation

The NIPG formulation will be derived under the assumption that ψ is a solution to the Poisson system and satisfies the regularity condition $\psi \in H^1(\Omega) \cap H^2(\mathcal{T}_h)$. The main implications of this condition is that for any interior face, we have that both $[\,\psi\,] = 0$ and $[\,\nabla\psi\,] = 0$ *a.e.* Even though the final weak formulation is derived by assuming the above regularity condtion, we will see that this condition can be relaxed without the final formulation becoming ill-defined.

To derive the NIPG formulation, we begin by multiplying (3.1) by an arbitrary test function $\theta \in H^1(\mathcal{T}_h)$, and then integrating the resulting expression over an arbitrary element $K_j \in \mathcal{T}_h$. This leads to the local equation

$$\int_{K_j} (A\nabla\psi)\cdot\nabla\theta\,dx \;-\; \int_{\partial K_j} (A\nabla\psi)_{|K_j}\cdot\nu_{K_j}\,\theta_{|K_j}\,dS \;=\; \int_{K_j} F\,\theta\,dx\,. \tag{3.4}$$

Performing the same calculation for every element in the mesh and then summing each of

the resulting local equations leads to the equation

$$\sum_{j=1}^{N_h} \int_{K_j} (A\nabla\psi)\cdot\nabla\theta\,dx \;-\; \sum_{K_j\in\mathcal{T}_h} \int_{\partial K_j} (A\nabla\psi)_{|_{K_j}}\cdot\nu_{_{K_j}}\,\theta_{|_{K_j}}\,dS \;=\; \int_\Omega F\,\theta\,dx\,. \tag{3.5}$$

Lemma 5 is now applied to the lefthandside term in (3.5) that contains integrations on $\partial K_{_j}$. The resulting expression is then decomposed into those integrations on interior faces and boundary faces. The boundary face integrations are further decomposed to those on the Dirichlet boundary and those on the Neumann boundary. We then weakly impose the Neumann boundary condition (3.3). After performing these calculations, (3.5) becomes

$$\begin{aligned}\sum_{j=1}^{N_h} \int_{K_j} (A\nabla\psi)\cdot\nabla\theta\,dx \;-\; \sum_{k=1}^{P_h} \int_{f_k} \left(\{\,A\nabla\psi\cdot\nu_{_k}\,\}\,[\,\theta\,] \;+\; [\,A\nabla\psi\cdot\nu_{_k}\,]\,\{\,\theta\,\}\right) dS \\ -\; \sum_{f_k\in\Gamma_{_D}} \int_{f_k} (A\nabla\psi)\cdot\nu_{_k}\,\theta\,dS \;=\; \int_\Omega F\,\theta\,dx \;+\; \sum_{f_k\in\Gamma_{_N}} \int_{f_k} r_{_N}\,\theta\,dS\,.\end{aligned} \tag{3.6}$$

The regularity properties of the matrix A and the true solution ψ allow for (3.6) to be simplified. Specifically, the facts $A\in C^1(\Omega)$ and $\psi\in H^1(\Omega)\times H^2(\mathcal{T}_h)$ together imply $[\,A\nabla\psi\cdot\nu_{_k}\,]=0$ *a.e.* on any interior face $f_{_k}$. Therefore, (3.6) reduces to

$$\begin{aligned}\sum_{j=1}^{N_h} \int_{K_j} (A\nabla\psi)\cdot\nabla\theta\,dx \;-\; \sum_{k=1}^{P_h} \int_{f_k} \{A\nabla\psi\cdot\nu_{_k}\}\,[\,\theta\,]\,dS \;-\; \sum_{f_k\in\Gamma_{_D}} \int_{f_k} (A\nabla\psi)\cdot\nu_{_k}\,\theta\,dS \\ =\; \int_\Omega F\,\theta\,dx \;+\; \sum_{f_k\in\Gamma_{_N}} \int_{f_k} r_{_N}\,\theta\,dS\,.\end{aligned} \tag{3.7}$$

Antisymmetrization terms

We now turn attention to rewriting (3.7) in such a way as to help simplify the error analysis. The reason for doing this is that the final variational formulation will in large part determine the quantities in which we can find suitable error bounds. It is desirable to find a formulation such that the term $-\sum_{k=1}^{P_h}\int_{f_k}\{A\nabla\psi\cdot\nu_{_k}\}\,[\,\theta\,]\,dS$ in (3.7) is canceled when the test function θ is taken to be ψ. To this end, consider the term $\sum_{k=1}^{P_h}\int_{f_k}\{A\nabla\theta\cdot\nu_{_k}\}\,[\,\psi\,]\,dS$. This term is equal to zero, which follows by the regularity properties of ψ, since $[\,\psi\,]=0$ *a.e.* on any

interior face f_k. As mentioned, it also satisfies the cancelation property

$$-\sum_{k=1}^{P_h}\int_{f_k}\{A\nabla\psi\cdot\nu_k\}\,[\,\theta\,]\,dS\;+\;\sum_{k=1}^{P_h}\int_{f_k}\{A\nabla\theta\cdot\nu_k\}\,[\,\psi\,]\,dS\;=\;0\,, \tag{3.8}$$

when θ is set equal to ψ. We remark that if we think of the lefthandside in (3.8) as a bilinear operator B acting on $H^1(\mathcal{T}_h)\times H^1(\mathcal{T}_h)$, then this operator is nonsymmetric. In fact, it is antisymmetric, i.e., $B(\phi,\theta)=-B(\theta,\phi),\ \forall\phi,\theta\in H^1(\mathcal{T}_h)$. After adding the antisymmetrization term $\sum_{k=1}^{P_h}\int_{f_k}\{A\nabla\theta\cdot\nu_k\}\,[\,\psi\,]\,dS$ to the lefthandside of (3.7), we get the equation

$$\begin{aligned}\sum_{j=1}^{N_h}\int_{K_j}(A\nabla\psi)\cdot\nabla\theta\,dx\;-\;\sum_{k=1}^{P_h}\int_{f_k}\{A\nabla\psi\cdot\nu_k\}\,[\,\theta\,]\,dS\;+\;\sum_{k=1}^{P_h}\int_{f_k}\{A\nabla\theta\cdot\nu_k\}\,[\,\psi\,]\,dS\\ -\sum_{f_k\in\Gamma_D}\int_{f_k}(A\nabla\psi)\cdot\nu_k\,\theta\,dS\;=\;\int_\Omega F\,\theta\,dx\;+\;\sum_{f_k\in\Gamma_N}\int_{f_k}r_N\,\theta\,dS\,.\end{aligned} \tag{3.9}$$

Error analysis concerns also lead to us to consider finding a formulation such that the Dirichlet boundary term $-\sum_{f_k\in\Gamma_D}\int_{f_k}(A\nabla\psi)\cdot\nu_k\,\theta\,dS$ in (3.9) is canceled when $\theta=\psi$. This can be accomplished by adding the antisymmetrization term $\sum_{f_k\in\Gamma_D}\int_{f_k}(A\nabla\theta)\cdot\nu_k\,\psi\,dS$ to the lefthandside of (3.9). However, there is no reason that this term must be equal to zero. Therefore, we offset the addition of this term by adding $\sum_{f_k\in\Gamma_D}\int_{f_k}(A\nabla\theta)\cdot\nu_k\,r_D\,dS$ to the righthandside of (3.9), which follows since $\psi=r_D$ on Γ_D. Adding these two terms to the formulation yields the equation

$$\begin{aligned}\sum_{j=1}^{N_h}\int_{K_j}(A\nabla\psi)\cdot\nabla\theta\,dx\;-\;\sum_{k=1}^{P_h}\int_{f_k}\{A\nabla\psi\cdot\nu_k\}\,[\,\theta\,]\,dS\;+\;\sum_{k=1}^{P_h}\int_{f_k}\{A\nabla\theta\cdot\nu_k\}\,[\,\psi\,]\,dS\\ -\sum_{f_k\in\Gamma_D}\int_{f_k}(A\nabla\psi)\cdot\nu_k\,\theta\,dS\;+\;\sum_{f_k\in\Gamma_D}\int_{f_k}(A\nabla\theta)\cdot\nu_k\,\psi\,dS\\ =\;\int_\Omega F\,\theta\,dx\;+\;\sum_{f_k\in\Gamma_N}\int_{f_k}r_N\,\theta\,dS\;+\;\sum_{f_k\in\Gamma_D}\int_{f_k}(A\nabla\theta)\cdot\nu_k\,r_D\,dS\,.\end{aligned} \tag{3.10}$$

The utility of the antisymmetrization terms will be fully appreciated in the discussion of the error analysis. In particular, it will be seen that these terms play a fundamental role in determining exactly what quantities can be estimated.

Penalty terms

At this point, the formulation is nearly complete. What remains to be finished is to find a way to ensure that if there are solutions to the true and discrete NIPG formulations, then these solutions are in fact unique. In NIPG, this is done through the use of penalty terms.

The first penalty term we consider is the following symmetric interior penalty term along the interior faces:

$$\sum_{k=1}^{P_h} \frac{r\sigma_k}{|\, f_k \,|^\beta} \int_{f_k} [\, \psi \,]\, [\, \theta \,]\, dx\,, \tag{3.11}$$

where the paremeters σ_k are each fixed constants satisfying $0 < \sigma_0 < \sigma_k < \sigma_m$, and the parameter $\beta \geq 1/2$ is a fixed constant. By the assumed regularity properties of ψ, we see that (3.11) is equal to zero. The penalty parameter $\frac{r\sigma_k}{|\, f_k \,|^\beta}$ is seen to increase both as the polynomial order of the discrete approximation space increases and as the area of the interior faces decreases. This property of the penalty parameter is desirable, since it reflects the fact that the approximate solution to ψ should have better regularity across the interior faces as h and r are refined.

As for the Dirichlet boundary faces, we consider the symmetric boundary penalty term

$$\sum_{f_k \in \Gamma_D} \frac{r\sigma_k}{|\, f_k \,|^\beta} \int_{f_k} \psi\, \theta\, dx\,. \tag{3.12}$$

This term is not known to be zero, so it must be offset by the term

$$\sum_{f_k \in \Gamma_D} \frac{r\sigma_k}{|\, f_k \,|^\beta} \int_{f_k} r_{_D}\, \theta\, dx\,. \tag{3.13}$$

The formulation is now completed by adding (3.11) and (3.12) to the lefthandside of (3.10) and by adding (3.13) to the righthandside of (3.10). Adding these three terms results in

the equation

$$\sum_{j=1}^{N_h} \int_{K_j} (A\nabla\psi)\cdot\nabla\theta\, dx \;-\; \sum_{k=1}^{P_h} \int_{f_k} \{A\nabla\psi\cdot\nu_k\}\,[\,\theta\,]\, dS \;+\; \sum_{k=1}^{P_h} \int_{f_k} \{A\nabla\theta\cdot\nu_k\}\,[\,\psi\,]\, dS$$
$$- \sum_{f_k\in\Gamma_D} \int_{f_k} (A\nabla\psi)\cdot\nu_k\,\theta\, dS \;+\; \sum_{f_k\in\Gamma_D} \int_{f_k} (A\nabla\theta)\cdot\nu_k\,\psi\, dS$$
$$+ \sum_{k=1}^{P_h} \frac{r\sigma_k}{|\,f_k\,|^\beta} \int_{f_k} [\,\psi\,]\,[\,\theta\,]\, dx \;+\; \sum_{f_k\in\Gamma_D} \frac{r\sigma_k}{|\,f_k\,|^\beta} \int_{f_k} \psi\,\theta\, dS \;=\; \int_\Omega F\,\theta\, dx$$
$$+ \sum_{f_k\in\Gamma_N} \int_{f_k} r_{_N}\,\theta\, dS \;+\; \sum_{f_k\in\Gamma_D} \int_{f_k} (A\nabla\theta)\cdot\nu_k\, r_{_D}\, dS \;+\; \sum_{f_k\in\Gamma_D} \frac{r\sigma_k}{|\,f_k\,|^\beta} \int_{f_k} r_{_D}\theta\, dS\,. \tag{3.14}$$

Another important benefit of adding the Dirichlet penalty term is that the above formulation now weakly imposes the Dirichlet boundary condition. By both the Dirichlet and Neumann boundary conditions being weakly imposed in the (3.14), we avoid having to use a discrete approximation space whose member functions must satisfy the given boundary conditions. This is very advantageous when designing a computer code to implement the NIPG method, since one need not be concerned with finding appropiate conditions to be enforced in the code that ensure the computed solution behaves correctly on boundary domain $\partial\Omega$.

3.2.2 Weak problem statement

The formulation for the NIPG method is now complete. In order to simplify the notation, three functionals are introduced. First, the bilinear functional

$$a : (\,\psi,\theta\,) \in H^1(\mathcal{T}_h)\times H^1(\mathcal{T}_h) \to \mathbb{R}$$

is defined to be the first five terms of the lefthandside in (3.14), *i.e.*,

$$a(\,\psi,\theta\,) \;=\; \sum_{j=1}^{N_h} \int_{K_j} (A\nabla\psi)\cdot\nabla\theta\, dx - \sum_{k=1}^{P_h} \int_{f_k} \{A\nabla\psi\cdot\nu_k\}\,[\,\theta\,]\, dS \;+\; \sum_{k=1}^{P_h} \int_{f_k} \{A\nabla\theta\cdot\nu_k\}\,[\,\psi\,]\, dS$$
$$- \sum_{f_k\in\Gamma_D} \int_{f_k} (A\nabla\psi)\cdot\nu_k\,\theta\, dS \;+\; \sum_{f_k\in\Gamma_D} \int_{f_k} (A\nabla\theta)\cdot\nu_k\,\psi\, dS\,. \tag{3.15}$$

Since this functional contains both of the antisymmetrization terms, it is not a symmetric operator.

As previously mentioned, the cancelation of properties of the anti-symmetrization terms in the bilinear functional a allow has important implications when analyzing the NIPG method. In particular, we have the following crucial property:

Property 1. *For any arbitrary* $\theta \in H^1(\mathcal{T}_h)$, $a(\,\theta,\theta\,)$ *satisfies*

$$a(\,\theta,\theta\,) \;=\; \|\!|\, A^{1/2}\,\nabla\theta\,|\!\|^2_{0,\Omega}\,. \tag{3.16}$$

This property implies that $a(\,\theta,\theta\,)$ is a positive semi-definite function. However, it is not positive-definite since $a(\,\theta,\theta\,)$ equals the summation of terms involving only the local gradients of θ, and not θ itself.

Second, we introduce the symmetric, bilinear, penalty functional

$$J : (\,\psi,\theta\,) \in H^1(\mathcal{T}_h) \times H^1(\mathcal{T}_h) \to \mathbb{R}\,,$$

which is defined to be the last two terms of the lefthandside in (3.14), *i.e.*,

$$\begin{aligned} J(\,\psi,\theta\,) \;&=\; \sum_{k=1}^{P_h} \frac{r\sigma_k}{|\,f_k\,|^\beta} \int_{f_k} [\,\psi\,]\,[\,\theta\,]\,dx \;+\; \sum_{f_k\in\Gamma_{_D}} \frac{r\sigma_k}{|\,f_k\,|^\beta} \int_{f_k} \psi\,\theta\,dS\;. \\ &=\; J_{_{int}}(\,\psi,\theta\,) \;+\; J_{_D}(\,\psi,\theta\,), \end{aligned} \tag{3.17}$$

where $J_{_{int}}$ is interior penalty bilinear functional and $J_{_D}$ is the Dirichlet boundary bilinear penalty functional.

The symmetric nature of J yields the following important property:

Property 2. *For any arbitrary* $\theta \in H^1(\mathcal{T}_h)$, $J(\,\theta,\theta\,)$ *satisfies*

$$J(\,\theta,\theta\,) \;=\; \sum_{k=1}^{P_h} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,[\,\theta\,]\,\|^2_{0,f_k} \;+\; \sum_{f_k\in\Gamma_{_D}} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\theta\,\|^2_{0,f_k}\;, \tag{3.18}$$

which implies that $J(\,\theta,\theta\,)$ is a positive semi-definite function. However, it is not positive-definite since it only involves the jumps across the interior faces and the Dirichlet boundary values of a given function, but gives no information concerning the values of the function in the interiors of the elements.

The power of the NIPG formulation is understood when one considers the operator $a + J$.

In particular, we have that, $\forall\, \theta \in H^1(\mathcal{T}_h)$,

$$a(\,\theta,\theta\,) + J(\,\theta,\theta\,) \;=\; \|\!|\, A^{1/2}\,\nabla\theta\,|\!\|^2_{0,\Omega} \;+\; \sum_{k=1}^{P_h} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,[\,\theta\,]\,\|^2_{0,f_k} \;+\; \sum_{f_k\in\Gamma_{_D}} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\theta\,\|^2_{0,f_k}\,. \tag{3.19}$$

The above inequality can be used to show that $a(\,\theta,\theta\,) + J(\,\theta,\theta\,)$ is a positive-definite function, since a controls the values of $\nabla\theta$ in the element interiors and J controls the values of θ across the interior faces and on the Dirichlet boundary. Moreover, if we define the function $\|\cdot\|_{NIPG} : H^1(\mathcal{T}_h) \to \mathbb{R}$ by

$$\|\,\theta\,\|^2_{NIPG} \;=\; a(\,\theta,\theta\,) + J(\,\theta,\theta\,)\,, \qquad \theta \in H^1(\mathcal{T}_h)\,, \tag{3.20}$$

then this function is in fact a norm on $H^1(\mathcal{T}_h)$, provided that $|\Gamma_{_D}| > 0$.

Lemma 10. *The function $\|\cdot\|_{NIPG}$ is a norm on $H^1(\mathcal{T}_h)$.*

Proof. Assume that $\theta \in H^1(\mathcal{T}_h)$ and $\|\,\theta\,\|_{NIPG} = 0$. It then follows from (3.19) that $\nabla\theta = 0$ in the interior of each element, which implies that θ is piecewise constant function on $\mathcal{T}_h$. However, (3.19) also implies that the jump values of θ across interior faces are all equal to zero. So, θ must constant on Ω. Since $|\Gamma_{_D}| > 0$ is assumed, (3.19) implies that θ is zero on $\Gamma_{_D}$. Combining this with fact that θ is constant on Ω implies that $\theta \equiv 0$, in $L^2(\Omega)$.

For any arbitrary θ, $\phi \in H^1(\mathcal{T}_h)$, we have that

$$\begin{aligned}
\|\,\theta \;+\; \phi\,\|^2_{NIPG} \;&=\; \|\!|\, A^{1/2}\,\nabla\theta + A^{1/2}\nabla\phi\,|\!\|^2_{0,\Omega} \;+\; \sum_{k=1}^{P_h} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,[\,\theta\,] + [\,\phi\,]\,\|^2_{0,f_k} \\
&+\; \sum_{f_k\in\Gamma_{_D}} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\theta+\phi\,\|^2_{0,f_k} \;\le\; \Big(\,\|\!|\, A^{1/2}\,\nabla\theta\,|\!\|_{0,\Omega} \;+\; \|\!|\, A^{1/2}\,\nabla\phi\,|\!\|_{0,\Omega}\,\Big)^2 \\
&+\; \sum_{k=1}^{P_h} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\Big(\,\|\,\theta\,\|_{0,f_k} \;+\; \|\,\phi\,\|_{0,f_k}\,\Big)^2 \;+\; \sum_{f_k\in\Gamma_{_D}} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\Big(\,\|\,\theta\,\|_{0,f_k} \;+\; \|\,\phi\,\|_{0,f_k}\,\Big)^2 \\
&\le\; \|\,\theta\,\|^2_{NIPG} \;+\; \|\,\phi\,\|^2_{NIPG} \;+\; 2\,\big(\,\|\!|\, A^{1/2}\,\nabla\theta\,|\!\|_{0,\Omega}\,\|\!|\, A^{1/2}\,\nabla\phi\,|\!\|_{0,\Omega} \\
&+\; \sum_{k=1}^{P_h} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\theta\,\|_{0,f_k} \sum_{k=1}^{P_h} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\phi\,\|_{0,f_k} \;+\; \sum_{f_k\in\Gamma_{_D}} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\theta\,\|_{0,f_k} \sum_{f_k\in\Gamma_{_D}} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\phi\,\|_{0,f_k}\big) \\
&=\; \|\,\theta\,\|^2_{NIPG} \;+\; \|\,\phi\,\|^2_{NIPG} \;+\; 2\,(\,a_1\,b_1 \;+\; a_2\,b_2 \;+\; a_3\,b_3\,)^2\,,
\end{aligned} \tag{3.21}$$

where the constants a_1, a_2, a_3, b_1, b_2, b_3 are defined as

$$a_1 \;=\; \| A^{1/2}\,\nabla\theta \|_{0,\Omega}\,,\;\; a_2 \;=\; \sum_{k=1}^{P_h} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\theta\,\|_{0,f_k}\,,\;\; a_3 \;=\; \sum_{f_k\in\Gamma_D} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\theta\,\|_{0,f_k}\,,\;\text{and}$$

$$b_1 \;=\; \| A^{1/2}\,\nabla\phi \|_{0,\Omega}\,,\;\; b_2 \;=\; \sum_{k=1}^{P_h} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\phi\,\|_{0,f_k}\,,\;\; b_3 \;=\; \sum_{f_k\in\Gamma_D} \frac{r\sigma_k}{|\,f_k\,|^\beta}\,\|\,\phi\,\|_{0,f_k}\,.$$

Clearly, it follows that

$$\begin{aligned}
(\,a_1\,b_1 \;+\; a_2\,b_2 \;+\; a_3\,b_3\,)^2 \;&=\; a_1^2\,b_1^2 \;+\; a_2^2\,b_2^2 \;+\; a_3^2\,b_3^2 \\
&\quad +\; 2\,a_1\,b_1\,a_2\,b_2 \;+\; 2\,a_1\,b_1\,a_3\,b_3 \;+\; 2\,a_2\,b_2\,a_3\,b_3 \\
&\le\; a_1^2\,b_1^2 \;+\; a_2^2\,b_2^2 \;+\; a_3^2\,b_3^2 \\
&\quad +\; a_1^2\,b_2^2 \;+\; a_1^2\,b_3^2 \;+\; a_2^2\,b_1^2 \;+\; a_2^2\,b_3^2 \;+\; a_3^2\,b_1^2 \;+\; a_3^2\,b_2^2 \\
&=\; \left(\,a_1^2 \;+\; a_2^2 \;+\; a_3^2\,\right)\left(\,b_1^2 \;+\; b_2^2 \;+\; b_3^2\,\right) \\
&=\; \|\,\theta\,\|_{NIPG}^2\,\|\,\phi\,\|_{NIPG}^2\,. \qquad (3.22)
\end{aligned}$$

This then implies that

$$\begin{aligned}
\|\,\theta \;+\; \phi\,\|_{NIPG}^2 \;&\le\; \|\,\theta\,\|_{NIPG}^2 \;+\; \|\,\phi\,\|_{NIPG}^2 \;+\; 2\,\|\,\theta\,\|_{NIPG}^2\,\|\,\phi\,\|_{NIPG}^2 \\
&=\; (\,\|\,\theta\,\|_{NIPG} \;+\; \|\,\phi\,\|_{NIPG}\,)^2\,. \qquad (3.23)
\end{aligned}$$

So, $\|\cdot\|_{NIPG}$ satisfies the triangle inequality.

It is trivial to check that $\|\,\lambda\,\theta\,\|_{NIPG} \;=\; |\,\lambda\,|\;\|\,\theta\,\|_{NIPG}$ is satified, $\forall\,\lambda\in\mathbb{R}$, $\forall\,\theta\in H^1(\mathcal{T}_h)$. □

Lastly, the linear functional $L : \theta \in H^1(\mathcal{T}_h) \to \mathbb{R}$ is defined to be the righthandside in (3.14), *i.e.*,

$$\begin{aligned}
L(\,\theta\,) \;&=\; \int_\Omega F\,\theta\,dx \;+\; \sum_{f_k\in\Gamma_N} \int_{f_k} r_{_N}\,\theta\,dS \;+\; \sum_{f_k\in\Gamma_D} \int_{f_k} (A\nabla\theta)\cdot\nu_k\,r_{_D}\,dS \\
&\quad +\; \sum_{f_k\in\Gamma_D} \frac{r\sigma_k}{|\,f_k\,|^\beta} \int_{f_k} r_{_D}\theta\,dS\,. \qquad (3.24)
\end{aligned}$$

By design, L weakly imposes both the Dirichlet and Neumann boundary conditions and is independent of true solution ψ.

Using the above functional notation in conjunction with (3.14), the definition of what it

means to be a true NIPG solution to the model problem is now easily defined as follows:

Definition 7. *A function $\psi \in H^1(\mathcal{T}_h)$ is said to be a true NIPG solution of the Poisson system (3.1)-(3.3) if*

$$a(\,\psi,\theta\,) \;+\; J(\,\psi,\theta\,) \;=\; L(\,\theta\,)\,, \tag{3.25}$$

is satisfied, $\forall\,\theta \in H^1(\mathcal{T}_h)$.

We note that (3.25) is linear in both ψ and θ.

The uniqueness property of the NIPG method is now easily established by taking of advantage of the linearity of (3.25) and the fact that $\|\cdot\|_{NIPG}$ is a norm on $H^1(\mathcal{T}_h)$.

Lemma 11. [Uniqueness of the true NIPG method] *If ψ is a true NIPG solution to the Poisson system, then it is unique.*

Proof. Suppose that ψ and ϕ are both true NIPG solutions to the Poisson system. Since

$$\begin{aligned} a(\,\psi,\theta\,) \;+\; J(\,\psi,\theta\,) \;&=\; L(\,\theta\,) \qquad \text{and} \\ a(\,\phi,\theta\,) \;+\; J(\,\phi,\theta\,) \;&=\; L(\,\theta\,)\,, \end{aligned}$$

$\forall\,\theta \in H^1(\mathcal{T}_h)$, it follows that

$$a(\,\psi-\phi,\theta\,) \;+\; J(\,\psi-\phi,\theta\,) \;=\; 0\,,$$

$\forall\,\theta \in H^1(\mathcal{T}_h)$. Upon setting $\theta = \psi - \phi$, we get that

$$\|\,\psi-\phi\,\|_{NIPG} \;=\; 0\,.$$

Hence, $\psi \equiv \phi$. □

The consistency of NIPG method, with respect to the distributional solution from Theorem 10, is now stated without proof.

Lemma 12. [Consistency of the NIPG method] *If the distributional solution ψ given in Theorem 10 satisfies the additional regularity condition $\psi \in H^2(\mathcal{T}_h)$, then ψ is also a unique, true NIPG solution to the Poisson system. Conversely, if ψ is a true NIPG solution to the Poisson system, and hence unique, and satisfies the additional regularity condition $\psi \in H^2(\mathcal{T}_h)$, then ψ is also the unique distributional solution given in Theorem 10.*

The main idea behind the proof is that functions in the space $H^1(\Omega) \times H^2(\mathcal{T}_h)$ have zero jump values across the interior faces of the mesh. By using this fact, the above lemma is easy to prove.

The definition of what it means to be a discrete NIPG true solution of the Poisson system is the same as that which was given in the case for the true NIPG solution, except that the infinite dimensional space $H^1(\mathcal{T}_h)$ is substituted with the finite dimensional space $D^r(\mathcal{T}_h)$.

Definition 8. *A function $\psi_h \in D^r(\mathcal{T}_h)$ is said to be a discrete NIPG solution of the Poisson system (3.1)-(3.3) if*

$$a(\,\psi_h, \theta_h\,) \;+\; J(\,\psi_h, \theta_h\,) \;=\; L(\,\theta_h\,)\,, \tag{3.26}$$

is satisfied, $\forall\, \theta_h \in D^r(\mathcal{T}_h)$.

The NIPG problem is equivalent to solving a matrix equation, where the matrix, the so called stiffness matrix, is square. So, the existence and uniqueness of a discrete NIPG solution to the Poisson system is equivalent to the corresponding stiffness matrix being invertible. For the discrete problem, we have following:

Lemma 13. [Existence and uniqueness of the discrete NIPG method] *There exists a unique discrete NIPG solution $\psi_h \in D^r(\mathcal{T}_h)$ to the Poisson system.*

Proof. Using the identical proof that was used to prove uniqueness of solution for the true NIPG method, we get the uniqueness in the discrete case as well. Since existence and uniqueness are equivalent in the discrete case, the proof is complete. □

We now mention an indispensable property in the upcoming error analysis that follows from definitions 7 and 8. Suppose that ψ and ψ_h are true and discrete NIPG solutions to the Poisson system, respectively. Then from definitions 7 and 8, we get that

$$a(\psi_h, \theta_h) \;+\; J(\psi_h, \theta_h) \;=\; a(\psi, \theta_h) \;+\; J(\psi, \theta_h)\,, \tag{3.27}$$

is satisfied, $\forall\, \theta_h \in D_r(\mathcal{T}_h)$. This equality can be rewritten as

$$a(\psi - \psi_h, \theta_h) \;+\; J(\psi - \psi_h, \theta_h) \;=\; 0\,, \tag{3.28}$$

$\forall\, \theta_h \in D_r(\mathcal{T}_h)$. This property is known as the Galerkin orthogonality property. The way this relationship enters into the error analysis will be seen shortly.

3.2.3 *A priori* error estimate

We now review the original error estimate for the NIPG method first given in [68]. This will be carried out by first looking at the bilinear functionals a and J to determine what quantities involving $\psi - \psi_h$ can be estimated in a hp sense. Next, it will be shown how the problem of finding the error between ψ and ψ_h can be decomposed in two seperate problems, one of finding the error between ψ and $\Pi_h^r \psi$, where $\Pi_h^r \psi$ satisfies properties (2.94)-(2.96),and one of finding the error between ψ_h and $\Pi_h^r \psi$. The remainder of this subsection will then give a rough sketch of how the quantities depending on $\psi - \Pi_h^r \psi$ are estimated and how those that depend on $\psi_h - \Pi_h^r \psi$ are estimated.

To see what quantities that depend on the difference between ψ and ψ_h the NIPG method allows one to bound, we set $\theta = \psi - \psi_h \in H^1(\mathcal{T}_h)$ in (3.19) to get that

$$\begin{aligned} \| \psi \ - \ \psi_h \|_{NIPG}^2 \ &= \ \| A^{1/2} \nabla(\psi - \psi_h) \|_{0,\Omega}^2 \ + \ \sum_{k=1}^{P_h} \frac{r\sigma_k}{| f_k |^\beta} \| [\psi - \psi_h] \|_{0,f_k}^2 \\ &+ \ \sum_{f_k \in \Gamma_D} \frac{r\sigma_k}{| f_k |^\beta} \| \psi - \psi_h \|_{0,f_k}^2 \, . \end{aligned} \tag{3.29}$$

The three terms in the righthandside of (3.29) give a measure of the errors, with respect to the L^2-norm, between $A^{1/2}\nabla\psi$ and $A^{1/2}\nabla\psi_h$ over the domain Ω, ψ and ψ_h across the interior faces, and ψ and ψ_h on the Dirichelt boundary faces, respectively.

The error between $\nabla\psi$ and $\nabla\psi_h$ can be obtained in terms of the error between $A^{1/2}\nabla\psi$ and $A^{1/2}\nabla\psi_h$ from the inequality

$$\| \nabla\psi - \nabla\psi_h \|_{0,\Omega} \ \leq \ C \| A^{1/2}(\nabla\psi - \nabla\psi_h) \|_{0,\Omega} \, , \tag{3.30}$$

which follows from the fact that the matrix function A is symmetric and uniformly positive definite. In particular, these two matrix properties together imply that the Rayleigh quotient of A is uniformly bounded above and below, i.e., there exist fixed constants $0 < \lambda_{min} < \lambda_{max}$ such that

$$\forall \, \xi \in \mathbb{R}^3, \ \forall \, x \in \Omega, \qquad \lambda_{min} \| \xi \|^2 \ \leq \ \| \xi^T A(x) \, \xi \| \ \leq \ \lambda_{max} \| \xi \|^2 \, . \tag{3.31}$$

It is easy to see that (3.30) follows immediately from (3.31), with $C = \lambda_{min}^{-1}$.

In order to begin the estimation problem, we first decompose $\| \psi - \psi_h \|_{NIPG}$ in (3.29) as

follows:

$$\begin{aligned} \| \psi \ - \ \psi_h \|^2_{NIPG} \ &= \ \| \psi \ - \ \Pi^r_h \psi \ + \ \Pi^r_h \psi \ - \ \psi_h \|^2_{NIPG} \\ &\leq \ (\| \psi \ - \ \Pi^r_h \psi \|_{NIPG} \ + \ \| \psi_h \ - \ \Pi^r_h \psi \|_{NIPG})^2 \\ &= \ 2 \| \psi \ - \ \Pi^r_h \psi \|^2_{NIPG} \ + \ 2 \| \psi_h \ - \ \Pi^r_h \psi \|^2_{NIPG} . \end{aligned} \tag{3.32}$$

This inequality shows that the error between ψ and ψ_h can be estimated by bounding the error between ψ and $\Pi^r_h \psi$ and the error between ψ_h and $\Pi^r_h \psi$.

To estimate $\| \psi - \Pi^r_h \psi \|_{NIPG}$, we use the interpolation properties (2.94)-(2.95) of $\Pi^r_h \psi$ and the property (3.31) of the matrix A to bound each of the three seperate terms compromising this norm.

Using the matrix property (3.31) and then (2.94), a bound for the first term in $\| \psi - \Pi^r_h \psi \|_{NIPG}$ is gotten by

$$\| A^{1/2} \nabla (\psi - \Pi^r_h \psi) \|^2_{0,\Omega} \ \leq \ C \| \nabla (\psi - \Pi^r_h \psi) \|^2_{0,\Omega} \ \leq \ C \frac{h^{2\mu-2}}{r^{2s-2}} \| \psi \|^2_{s,\Omega} , \tag{3.33}$$

where $C = \lambda_{max}$.

To find a suitable bound for the second term, we write

$$\begin{aligned} &\sum_{k=1}^{P_h} \frac{r\sigma_k}{| f_k |^\beta} \| [\psi - \Pi^r_h \psi] \|^2_{0,f_k} \ \leq \ C r h^{-2\beta} \sum_{k=1}^{P_h} \| (\psi - \Pi^r_h \psi)_{|K_1} \ - \ (\psi - \Pi^r_h \psi)_{|K_2} \|^2_{0,f_k} \\ &\leq \ C r h^{-2\beta} \sum_{k=1}^{P_h} \left(\| (\psi - \Pi^r_h \psi)_{|K_1} \|_{0,f_k} \ + \ \| (\psi - \Pi^r_h \psi)_{|K_2} \|_{0,f_k} \right)^2 \\ &\leq \ C r h^{-2\beta} \sum_{k=1}^{P_h} \frac{h^{2\mu-1}}{r^{2s-1}} \left(\| \psi \|_{s,K_1} + \| \psi \|_{s,K_2} \right)^2 \ \leq \ C \frac{h^{2\mu-1-2\beta}}{r^{2s-2}} \sum_{k=1}^{P_h} \left(\| \psi \|^2_{s,K_1} + \| \psi \|^2_{s,K_2} \right) \\ &\leq \ C \frac{h^{2\mu-1-2\beta}}{r^{2s-2}} \| \psi \|^2_{s,\Omega} . \end{aligned} \tag{3.34}$$

In a manner similar to the way in which inequality (3.34) was derived, we get that

$$\sum_{k \in \Gamma_D} \frac{r\sigma_k}{| f_k |^\beta} \| [\psi - \Pi^r_h \psi] \|^2_{0,f_k} \ \leq \ C \frac{h^{2\mu-1-2\beta}}{r^{2s-2}} \| \psi \|^2_{s,\Omega} . \tag{3.35}$$

Upon combining (3.33)-(3.35), we get that

$$\| \psi \ - \ \Pi_h^r \psi \|_{NIPG}^2 \ \leq \ C \frac{h^{2\mu-2}}{r^{2s-2}} \| \psi \|_{s,\Omega}^2 \ + \ C \frac{h^{2\mu-1-2\beta}}{r^{2s-2}} \| \psi \|_{s,\Omega}^2 . \tag{3.36}$$

Since $\beta \geq \frac{1}{2}$, the final bound for (3.36) is of order $\min\{2\mu-2, 2\mu-1-2\beta\} \geq 2\mu-1-2\beta$ in h and $2s-2$ in r. Thus, in order to the obtain optimal convergence rate in h of $2\mu-2$, β is chosen to be $\frac{1}{2}$, which results in the bound

$$\| \psi - \Pi_h^r \psi \|_{NIPG}^2 \ \leq \ C \frac{h^{2\mu-2}}{r^{2s-2}} \| \psi \|_{s,\Omega}^2 . \tag{3.37}$$

To finish estimating the righthandside of (3.32), it remains to bound the quantity $\| \psi_h - \Pi_h^r \psi \|_{NIPG}$. For convenience, define $w_h = \psi_h - \Pi_h^r \psi$. Since $w \in D^r(\mathcal{T}_h)$, the Galerkin orthogonality property (3.28) implies that

$$\begin{aligned}
\| w_h \|_{NIPG}^2 \ &= \ a(\psi - \Pi_h^r \psi, w_h) + J(\psi - \Pi_h^r \psi, w_h) \\
&= \ \sum_{j=1}^{N_h} \int_{K_j} A\nabla(\psi - \Pi_h^r \psi) \cdot \nabla w_h \, dx \ - \ \sum_{k=1}^{P_h} \int_{f_k} \{ A\nabla(\psi - \Pi_h^r \psi) \cdot \nu_k \} [w_h] \, dS \\
&\quad + \ \sum_{k=1}^{P_h} \int_{f_k} \{ A\nabla w_h \cdot \nu_k \} [\psi - \Pi_h^r \psi] \, dS \ - \ \sum_{f_k \in \Gamma_D} \int_{f_k} A\nabla(\psi - \Pi_h^r \psi) \cdot \nu_k \, w_h \, dS \\
&\quad + \ \sum_{f_k \in \Gamma_D} \int_{f_k} (A\nabla w_h) \cdot \nu_k \, (\psi - \Pi_h^r \psi) \, dS \ + \ \sum_{k=1}^{P_h} \frac{r\sigma_k}{| f_k |^\beta} \int_{f_k} [\psi - \Pi_h^r \psi] [w_h] \, dS \\
&\quad + \ \sum_{f_k \in \Gamma_D} \frac{r\sigma_k}{| f_k |^\beta} \int_{f_k} (\psi - \Pi_h^r \psi) \, w_h \, dS . \\
&= \ E_1 \ + \ E_2 \ + \ E_3 \ + \ E_4 \ + \ E_5 \ + \ E_6 \ + \ E_7 .
\end{aligned} \tag{3.38}$$

In [68], *hp*-estimates were proved for each of the terms $E_1, \ldots, E_7$. It will be seen that these estimates are optimal with respect to h, but that they are suboptimal with respect to r.

The original estimates given are as follows:

$$E_1 \;\leq\; \frac{1}{6}\,|\!|\!|\, A^{1/2}\nabla w_h \,|\!|\!|^2_{0,\Omega} \;+\; C\,\frac{h^{2\mu-2}}{r^{2s-2}}\,|\!|\!|\,\psi\,|\!|\!|^2_{0,\Omega}\,, \tag{3.39}$$

$$E_2 \;\leq\; \frac{1}{4}\,J_{int}(\,w_h, w_h\,) \;+\; C\sum_{k=1}^{P_h} |\,f_k\,|^{\beta}\,\frac{\bar{h}_k^{2\mu-3}}{r^{2s-2}}\,|\!|\!|\,\psi\,|\!|\!|^2_{s,K_{12}}\,, \tag{3.40}$$

$$E_3 \;\leq\; \frac{1}{6}\,|\!|\!|\, A^{1/2}\nabla w_h \,|\!|\!|^2_{0,\Omega} \;+\; C\,\frac{h^{2\mu-2}}{r^{2s-3}}\,|\!|\!|\,\psi\,|\!|\!|^2_{s,\Omega}\,, \tag{3.41}$$

$$E_4 \;\leq\; \frac{1}{4}\,J_{D}(\,w_h, w_h\,) \;+\; C\sum_{f_k\in\Gamma_D} |\,f_k\,|^{\beta}\,\frac{h_{f_k}^{2\mu-3}}{r^{2s-2}}\,\|\,\psi\,\|^2_{s,K_{f_k}}\,, \tag{3.42}$$

$$E_5 \;\leq\; \frac{1}{6}\,|\!|\!|\, A^{1/2}\nabla w_h \,|\!|\!|^2_{0,\Omega} \;+\; C\,\frac{h^{2\mu-2}}{r^{2s-3}}\,|\!|\!|\,\psi\,|\!|\!|^2_{s,\Omega}\,, \tag{3.43}$$

$$E_6 \;\leq\; \frac{1}{4}\,J_{int}(\,w_h, w_h\,) \;+\; C\sum_{k=1}^{P_h} |\,f_k\,|^{-\beta}\,\frac{\bar{h}_k^{2\mu-1}}{r^{2s-2}}\,|\!|\!|\,\psi\,|\!|\!|^2_{s,K_{12}}\,, \tag{3.44}$$

$$E_7 \;\leq\; \frac{1}{4}\,J_{D}(\,w_h, w_h\,) \;+\; C\sum_{f_k\in\Gamma_D} |\,f_k\,|^{-\beta}\,\frac{h_{f_k}^{2\mu-1}}{r^{2s-2}}\,\|\,\psi\,\|^2_{s,K_{f_k}}\,. \tag{3.45}$$

We note that for a given boundary face f_k, K_{f_k} is used to denote the unique element satisfying $f_k \subset \partial K_{f_k}$.

Remark 3. *The estimates given in (3.39)-(3.45) all are of order $2s-2$ in r, except for those for E_3 and E_5, which are of order $2s-3$ in r. This shows that the estimates for these two terms account for the NIPG method being suboptimal in r.*

Upon combining results (3.39)-(3.45) and substituting them into the righthandside of (3.38), and then plugging in the definition of $\|\cdot\|_{NIPG}$ into the lefthandside of (3.38), we get that, after some algebraic manipulation,

$$\begin{aligned}
\|\,w_h\,\|^2_{NIPG} \;&\leq\; C\sum_{k=1}^{\mathbb{P}_h}\left(\,|\,f_k\,|^{\beta}\,\frac{\bar{h}_k^{2\mu-3}}{r^{2s-2}} \;+\; |\,f_k\,|^{-\beta}\frac{\bar{h}_k^{2\mu-1}}{r^{2s-2}}\right)|\!|\!|\,\psi\,|\!|\!|^2_{s,K_{12}} \\
&+\; C\sum_{f_k\in\Gamma_D}\left(\,|\,f_k\,|^{\beta}\,\frac{h_{f_k}^{2\mu-3}}{r^{2s-2}} + |\,f_k\,|^{-\beta}\frac{h_{f_k}^{2\mu-1}}{r^{2s-2}}\right)\|\,\psi\,\|^2_{s,K_{f_k}} \;+\; C\,\frac{h^{2\mu-2}}{r^{2s-2}}\,|\!|\!|\,\psi\,|\!|\!|^2_{s,\Omega} \;+\; C\,\frac{h^{2\mu-2}}{r^{2s-3}}\,|\!|\!|\,\psi\,|\!|\!|^2_{s,\Omega} \\
&\leq\; C\,\frac{h^{2\mu-3+2\beta}}{r^{2s-2}}\,|\!|\!|\,\psi\,|\!|\!|^2_{s,\Omega} \;+\; C\,\frac{h^{2\mu-1-2\beta}}{r^{2s-2}}\,|\!|\!|\,\psi\,|\!|\!|^2_{s,\Omega} \;+\; C\,\frac{h^{2\mu-2}}{r^{2s-2}}\,|\!|\!|\,\psi\,|\!|\!|^2_{s,\Omega} \;+\; C\,\frac{h^{2\mu-2}}{r^{2s-3}}\,|\!|\!|\,\psi\,|\!|\!|^2_{s,\Omega} \\
&\leq\; C\left(\frac{h^{2\mu-3+2\beta}}{r^{2s-2}} \;+\; \frac{h^{2\mu-1-2\beta}}{r^{2s-2}} \;+\; \frac{h^{2\mu-2}}{r^{2s-3}}\right)|\!|\!|\,\psi\,|\!|\!|^2_{s,\Omega}\,.
\end{aligned} \tag{3.46}$$

The final righthandside of the above inequality shows that $\| w_h \|^2_{NIPG}$ is of order $2s-3$ in r. However, its order with respect to h depends on the choice of the parameter β. It is easily seen that the order is at least $2s-2$, since one of the righthandside side terms contains the β-independent factor $h^{2\mu-2}$. If $\beta = 1/2$, then $2\mu-3+2\beta = 2\mu-2$ and $2\mu-1-2\beta = 2\mu-2$, which gives an overall order of $2\mu-2$ with respect to h for the estimate in (3.46). Any choice for β other than $\frac{1}{2}$ would result in an h-order that is less than the optimal order $2\mu-2$.

So, upon setting $\beta = 1/2$, (3.46) reduces to

$$\| w_h \|^2_{NIPG} \;\leq\; C\,\frac{h^{2\mu-2}}{r^{2s-3}}\,\| \psi \|^2_{s,\Omega}\,. \tag{3.47}$$

With the above estimates for $\| \psi - \Pi^r_h \psi \|^2_{NIPG}$ and $\| w_h \|^2_{NIPG}$ complete, we now state the original theorem given in [68] establishing an *a priori hp*-error estimate for the NIPG method.

Theorem 11. *Let ψ be the true NIPG solution and let ψ_h be the discrete NIPG solution. If $\psi \in H^s(\mathcal{T}_h)$, for $s \geq 1$, and $\beta = \frac{1}{2}$, then*

$$\| \psi \;-\; \psi_h \|^2_{NIPG} \;\leq\; C\,\frac{h^{2\mu-2}}{r^{2s-3}}\,\| \psi \|^2_{s,\Omega}\,. \tag{3.48}$$

Moreover, with this choice of β, the estimate (3.48) is optimal in h.

Proof. To obtain (3.48), simply plug in the estimates (3.37) and (3.47) into (3.32).

To show the optimality of (3.48) in h, we first see that it is easy to deduce the following:

$$\| \nabla(\psi - \psi_h) \|_{0,\Omega} \;\leq\; C\frac{h^{\mu-1}}{r^{s-\frac{3}{2}}}\,\| \psi \|_{s,\Omega}\,, \tag{3.49}$$

$$\| [\,\psi - \psi_h] \|_{0,\hat{\mathcal{F}}_h} \;\leq\; C\frac{h^{\mu-\frac{1}{2}}}{r^{s-1}}\,\| \psi \|_{s,\Omega}\,, \text{ and} \tag{3.50}$$

$$\| \psi - \psi_h \|_{0,f_k \in \Gamma_D} \;\leq\; C\frac{h^{\mu-\frac{1}{2}}}{r^{s-1}}\,\| \psi \|_{s,\Omega}\,. \tag{3.51}$$

We now find analogous estimates to (3.49)-(3.51) for the difference between the true solution ψ and its interpolant $\Pi^r_h \psi$. From (2.94), it directly follows that

$$\| \nabla(\,\psi - \Pi^r_h \psi\,) \|_{0,\Omega} \;\leq\; C\frac{h^{\mu-1}}{r^{s-1}}\,\| \psi \|_{s,\Omega}\,. \tag{3.52}$$

To obtain an estimate jump across the interior faces, we use the fact that for a given interior face $f_k = K_1 \cap K_2$ we have that

$$\begin{aligned} \| \psi - \Pi_h^r \psi \|_{0,f_k}^2 &\leq 2 \| (\psi - \Pi_h^r \psi)_{|K_1} \|_{0,f_k}^2 + 2 \| (\psi - \Pi_h^r \psi)_{|K_2} \|_{0,f_k}^2 \\ &\leq C \frac{h_j^{2\mu-1}}{r^{s-1}} \|\psi\|_{s,K_{12}}^2 , \end{aligned}$$

where the last line follows by (2.95). Hence,

$$\| [\psi - \Pi_h^r \psi] \|_{0,\mathring{\mathcal{F}}_h} \leq C \frac{h_j^{\mu-\frac{1}{2}}}{r^{s-\frac{1}{2}}} \| \psi \|_{s,\Omega} . \tag{3.53}$$

Along the Dirichlet boundary, (2.95) implies that

$$\| \psi - \Pi_h^r \psi \|_{0,f_k \in \Gamma_D} \leq C \frac{h^{\mu-\frac{1}{2}}}{r^{s-\frac{1}{2}}} \| \psi \|_{s,\Omega} . \tag{3.54}$$

Comparing the estimates (3.49)-(3.51) to the estimates (3.52)-(3.54), respectively, it is seen that (3.49)-(3.51) are optimal in h, since each of the three estimates achieves the same order of accuracy in h as does the corresponding estimate for the interpolant. □

Although the error estimate in Theorem 11 is optimal with respect to h, it is not true that it is optimal with respect to r. The reason is that inequalities (3.49)-(3.51) have orders of accuracy in r of $s-\frac{3}{2}$, $s-1$, and $s-1$, respectively. In contrast, the interpolant inequalities (3.52)-(3.54) have orders of accuracy in r of $s-1$, $s-\frac{1}{2}$, and $s-\frac{1}{2}$, respectively. Thus, the interpolant gives estimates that are better by a factor of $r^{-1/2}$ than the estimates for the NIPG approximation ψ_h.. This suggests that it might be possible to improve the NIPG estimates by a factor of $r^{-1/2}$. If such an improvement can be made, then the resulting estimates would be optimal not only in h, but also in r as well.

3.3 Improvement and extension of the *a priori* NIPG error estimate

Our goal now is to improve upon and expand the original NIPG error estimate. Specifically, we will show that it is possible to improve the NIPG error estimate in theorem 11 by a factor of $r^{-1/2}$. Also, error estimates will be obtained for the gradient of the NIPG approximation along the boundaries of the elements. The results presented are especially important for

problems where the gradient of the true solution ψ is of more interest than ψ itself. This is indeed the case for plasma systems modeled by the Vlasov-Poisson system, since the Vlasov equation depends directly on the electric field, which is the gradient of a potential that satisfies the Poisson equation. To our knowledge, all of the results presented in this section are new.

3.3.1 Improvement of the error estimate

As previously mentioned, the order of the error estimate in Theorem 11 with respect to r is limited by the estimates (3.41) and (3.43). If these estimates could be improved to an order of $2s-2$ in r, then the overall error estimate in Theorem 11 would be of the optimal order $2s-2$ as well. In [68], (3.41) and (3.43) were derived using a suboptimal inverse inequality for polynomial functions. This inverse inequality is similar to the inverse inequality (2.67) in Lemma 6. However, the result used in [68] has a righthandside bound that scales with a factor of r, whereas the result in Lemma 6 scales with the factor $r^{1/2}$. It is precisely by using the new results of Lemma 6 that an improved *a priori* NIPG error estimate that is optimal in both h and r can be proved.

The following theorem gives the improved NIPG error estimate.

Theorem 12. *Let ψ be the true NIPG solution and let ψ_h be the discrete NIPG solution. If $\psi \in H^s(\mathcal{T}_h)$, for $s \geq 1$, and $\beta = \frac{1}{2}$, then the following is satisfied:*

$$\begin{aligned} &\| A^{1/2} \nabla(\psi - \psi_h) \|_{0,\Omega}^2 \; + \; \sum_{k=1}^{P_h} \frac{r\sigma_k}{| f_k |^{\frac{1}{2}}} \| [\psi - \psi_h] \|_{0,f_k}^2 \; + \; \sum_{f_k \in \Gamma_D} \frac{r\sigma_k}{| f_k |^{\frac{1}{2}}} \| \psi - \psi_h \|_{0,f_k}^2 \\ &\leq \; C \frac{h^{2\mu-2}}{r^{2s-2}} \| \psi \|_{s,\Omega}^2 . \end{aligned} \tag{3.55}$$

Moreover, with this choice of β, the estimate (3.55) is optimal in both h and r.

Proof. The two quantities that scale with r^{2s-3} instead of r^{2s-2} in (3.39) are E_3 and E_5. If we can show both of these quantities indeed scale with r^{2s-2}, then (3.55) will be established. We recall that $w_h = \psi_h - \Pi_h^r \psi$.

We begin by proving a new bound for the E_3. By a simple application of the triangle

inequality we get that

$$E_3 \;=\; \sum_{k=1}^{P_h} \int_{f_k} |\{A\nabla w_h \cdot \nu_k\}\,[\,\psi - \Pi_h^r\,\psi\,]|\,dS \;\leq\; \sum_{k=1}^{P_h} \|\,\{A\nabla w_h \cdot \nu_k\}\,\|_{0,f_k}\,\|\,[\,\psi - \Pi_h^r\,\psi\,]\,\|_{0,f_k}\,. \tag{3.56}$$

Now, let $f_k = K_1 \cap K_2$ be an arbitrary interior face. Then

$$\begin{aligned}\|\,\{A\nabla w_h \cdot \nu_k\}\,\|_{0,f_k} \;&\leq \frac{1}{2}\|(A\nabla w_h)_{|K_1} \cdot \nu_k\|_{0,f_k} \;+\; \frac{1}{2}\|(A\nabla w_h)_{|K_2} \cdot \nu_k\|_{0,f_k} \\ &\leq C\,\|(A^{1/2}\nabla w_h)_{|K_1} \cdot \nu_k\|_{0,f_k} \;+\; C\,\|(A^{1/2}\nabla w_h)_{|K_2} \cdot \nu_k\|_{0,f_k}\,. \end{aligned} \tag{3.57}$$

This last line above follows the assumed properties of the matrix A.

We now make use the inverse inequality (2.68) from Lemma 6 by applying this result to both of the terms in the righthandside of (3.57). This leads to the inequality

$$\begin{aligned}\|\,\{A\nabla w_h \cdot \nu_k\}\,\|_{0,f_k} \;&\leq\; C\,r^{\frac{1}{2}}\,(\,h_{K_1}^{-\frac{1}{2}}\,\|A^{1/2}\nabla w_h\|_{0,K_1} \;+\; h_{K_2}^{-\frac{1}{2}}\,\|A^{1/2}\nabla w_h\|_{0,K_2}\,) \\ &\leq\; C\,\bar{h}_k^{-\frac{1}{2}} r^{\frac{1}{2}}\,(\,\|A^{1/2}\nabla w_h\|_{0,K_1} \;+\; \|A^{1/2}\nabla w_h\|_{0,K_2}\,) \\ &\leq\; C\,\bar{h}_k^{-\frac{1}{2}} r^{\frac{1}{2}}\,(\,\|A^{1/2}\nabla w_h\|_{0,K_1}^2 \;+\; \|A^{1/2}\nabla w_h\|_{0,K_2}^2\,)^{\frac{1}{2}} \\ &\leq\; C\,h^{-\frac{1}{2}} r^{\frac{1}{2}}\,|\!|\!|\,A^{1/2}\nabla w_h\,|\!|\!|_{0,K_{12}}\,. \end{aligned} \tag{3.58}$$

Using the *hp*-interpolation property (2.95), it follows that

$$\begin{aligned}\|\,[\,\psi - \Pi_h^r\,\psi\,]\,\|_{0,f_k} \;&\leq \|(\,\psi - \Pi_h^r\,\psi\,)_{|K_1}\,\|_{0,f_k} \;+\; \|(\,\psi - \Pi_h^r\,\psi\,)_{|K_2}\,\|_{0,f_k} \\ &\leq \frac{C}{r^{s-\frac{1}{2}}}\,(\,h_{K_1}^{\mu-\frac{1}{2}}\,\|\psi\|_{s,K_1} \;+\; h_{K_2}^{\mu-\frac{1}{2}}\,\|\psi\|_{s,K_2}\,) \\ &\leq C\,\frac{\bar{h}_k^{\mu-\frac{1}{2}}}{r^{s-\frac{1}{2}}}\,(\,\|\psi\|_{s,K_1}^2 \;+\; \|\psi\|_{s,K_2}^2\,)^{\frac{1}{2}} \;\leq\; C\,\frac{h^{\mu-\frac{1}{2}}}{r^{s-\frac{1}{2}}}\,|\!|\!|\,\psi\,|\!|\!|_{s,K_{12}}\,. \end{aligned} \tag{3.59}$$

Combining (3.58) and (3.59) with (3.57), we arrive at

$$
\begin{aligned}
E_3 \;&\leq\; \sum_{k=1}^{P_h} \| \{A\nabla w_h \cdot \nu_k\} \|_{0,f_k} \, \| [\, \psi - \Pi_h^r \psi \,] \|_{0,f_k} \\
&\leq\; \sum_{k=1}^{P_h} \left(C\, h^{-\frac{1}{2}} r^{\frac{1}{2}} \, |\!|\!| A^{1/2}\nabla w_h |\!|\!|_{0,K_{12}} \right) \left(\frac{h^{\mu-\frac{1}{2}}}{r^{s-\frac{1}{2}}} \, |\!|\!| \psi |\!|\!|_{s,K_{12}} \right) \\
&\leq\; C \left(\sum_{k=1}^{P_h} |\!|\!| A^{1/2}\nabla w_h |\!|\!|^2_{0,K_{12}} \right)^{1/2} \frac{h^{\mu-1}}{r^{s-1}} \left(\sum_{k=1}^{P_h} |\!|\!| \psi |\!|\!|^2_{s,K_{12}} \right)^{1/2} \\
&\leq\; C\, |\!|\!| A^{1/2}\nabla w_h |\!|\!|_{0,\Omega} \, \frac{h^{\mu-1}}{r^{s-1}} \, |\!|\!| \psi |\!|\!|_{s,\Omega} \\
&\leq\; \frac{1}{12} \, |\!|\!| A^{1/2}\nabla w_h |\!|\!|^2_{0,\Omega} \;+\; C\, \frac{h^{2\mu-2}}{r^{2s-2}} \, |\!|\!| \psi |\!|\!|^2_{s,\Omega} \,. \qquad (3.60)
\end{aligned}
$$

It remains to prove a new estimate for E_5. In the same manner in which (3.60) was derived, we proceed as follows:

$$
\begin{aligned}
E_5 \;&=\; \sum_{f_k \in \Gamma_D} \int_{f_k} |A\nabla w_h \cdot \nu_k \, (\, \psi - \Pi_h^r \psi \,)| \, dS \;\leq\; \sum_{f_k \in \Gamma_D} \| A\nabla w_h \cdot \nu_k \|_{0,f_k} \, \| \psi - \Pi_h^r \psi \|_{0,f_k} \\
&\leq\; C \sum_{f_k \in \Gamma_D} h_{f_k}^{-\frac{1}{2}} \, r^{\frac{1}{2}} \, \| A\nabla w_h \|_{0,K_{f_k}} \, \frac{h_{f_k}^{\mu-\frac{1}{2}}}{r^{s-\frac{1}{2}}} \, \| \psi \|_{s,K_{f_k}} \;\leq\; C \sum_{f_k \in \Gamma_D} \| A\nabla w_h \|_{0,K_{f_k}} \, \frac{h^{\mu-1}}{r^{s-1}} \, \| \psi \|_{s,K_{f_k}} \\
&\leq\; C \left(\sum_{f_k \in \Gamma_D} \| A\nabla w_h \|^2_{0,K_{f_k}} \right)^{1/2} \frac{h^{\mu-1}}{r^{s-1}} \left(\sum_{f_k \in \Gamma_D} \| \psi \|^2_{s,K_{f_k}} \right)^{1/2} \;\leq\; C\, |\!|\!| A\nabla w_h |\!|\!|_{0,\Omega} \, \frac{h^{\mu-1}}{r^{s-1}} \, |\!|\!| \psi |\!|\!|_{s,\Omega} \\
&\leq\; \frac{1}{6} \, |\!|\!| A^{1/2}\nabla w_h |\!|\!|^2_{0,\Omega} \;+\; C\, \frac{h^{2\mu-2}}{r^{2s-2}} \, |\!|\!| \psi |\!|\!|^2_{s,\Omega} \qquad (3.61)
\end{aligned}
$$

The estimates (3.60) and (3.61) are of order $2s-2$ in r. Upon replacing the suboptimal estimates (3.41) and (3.43) by (3.60) and (3.61), and then substituting these estimates into (3.38), and then substituting the resulting bound for $\| w_h \|^2_{NIPG}$ into (3.32), we get that estimate (3.55) is satisfied.

Using the improved *a priori* NIPG error estimate (3.55), it is easy to deduce the following

explicit error estimates:

$$\begin{aligned} \| \nabla(\psi - \psi_h) \|_{0,\Omega} &\leq C \frac{h^{\mu-1}}{r^{s-1}} \| \psi \|_{s,\Omega} , & (3.62) \\ \| [\psi - \psi_h] \|_{0,\mathring{\mathcal{F}}_h} &\leq C \frac{h^{\mu-\frac{1}{2}}}{r^{s-\frac{1}{2}}} \| \psi \|_{s,\Omega} , \text{ and} & (3.63) \\ \| \psi - \psi_h \|_{0,f_k \in \Gamma_D} &\leq C \frac{h^{\mu-\frac{1}{2}}}{r^{s-\frac{1}{2}}} \| \psi \|_{s,\Omega} . & (3.64) \end{aligned}$$

Comparing the above estimates for the NIPG approximation ψ_h with those in (3.52)-(3.54) for the interpolant $\Pi_h^r \psi$, it follows that (3.62)-(3.64) are each optimal in both h and r. This implies that (3.55) is optimal as well. □

3.3.2 Extension of the error estimate

The original NIPG estimate given in Theorem 11 and its improved version given in Theorem 12 both give error estimates for three quantities: the gradient of $A(\psi - \psi_h)$ over each element K_j, the jump of $\psi - \psi_h$ across all of the interior faces, and for $\psi - \psi_h$ on the Dirichlet boundary. However, in plasma systems it is the gradient of potential that is needed in the Vlasov equation. In particular, it is desirable to have as many estimates as possible for the local gradients of $\psi - \psi_h$. In particular, we would like to have optimal estimates for the quantities $\| [\nabla\psi - \nabla\psi_h] \|_{0,\mathring{\mathcal{F}}_h}$ and $\| \{\nabla\psi - \nabla\psi_h\} \|_{0,\mathring{\mathcal{F}}_h}$. The following lemma establishes estimates for these two quantitities of interest.

Lemma 14. *Let ψ be the true NIPG solution and let ψ_h be the discrete NIPG solution. If $\psi \in H^s(\mathcal{T}_h)$, for $s > 3/2$, and $\beta = \frac{1}{2}$, then the following are satisfied:*

$$\begin{aligned} \| [\nabla(\psi - \psi_h)] \|_{0,\mathring{\mathcal{F}}_h} &\leq C \frac{h^{\mu-\frac{3}{2}}}{r^{s-\frac{3}{2}}} \| \psi \|_{s,\Omega} \qquad \text{and} & (3.65) \\ \| \{\nabla(\psi - \psi_h)\} \|_{0,\mathring{\mathcal{F}}_h} &\leq C \frac{h^{\mu-\frac{3}{2}}}{r^{s-\frac{3}{2}}} \| \psi \|_{s,\Omega} . & (3.66) \end{aligned}$$

Moreover, with this choice of β, the estimate (3.55) is optimal in both h and r.

Proof. We first consider (3.65). Let $f_k = K_1 \cap K_2$ be an arbitray interior face. Then

$$\begin{aligned}
\| [\nabla\psi - \nabla\psi_h] \|^2_{0,f_k} \;&\leq\; \big(\| (\nabla\psi - \Pi^r_h \psi)_{|K_1} \|_{0,f_k} \;+\; \| (\nabla\psi - \Pi^r_h \psi)_{|K_2} \|_{0,f_k} \\
&\quad +\; \| (\nabla\psi_h - \Pi^r_h \psi)_{|K_1} \|_{0,f_k} \;+\; \| (\nabla\psi_h - \Pi^r_h \psi)_{|K_2} \|_{0,f_k} \big)^2 \\
&\leq\; 4 \| (\nabla\psi - \Pi^r_h \psi)_{|K_1} \|^2_{0,f_k} \;+\; 4 \| (\nabla\psi - \Pi^r_h \psi)_{|K_2} \|^2_{0,f_k} \\
&\quad +\; 4 \| (\nabla\psi_h - \Pi^r_h \psi)_{|K_1} \|^2_{0,f_k} \;+\; 4 \| (\nabla\psi_h - \Pi^r_h \psi)_{|K_2} \|^2_{0,f_k} \,. \qquad (3.67)
\end{aligned}$$

We now apply the interpolation estimate (2.96) to the terms involving $\Pi^r_h \psi$ and we apply the inverse inequality (2.89) to the terms involving ψ_h in the above inequality to get that

$$\begin{aligned}
\| [\nabla\psi - \nabla\psi_h] \|^2_{0,f_k} \;&\leq\; C \frac{h^{2\mu-3}_{K_1}}{r^{2s-3}} \|\psi\|^2_{s,K_1} \;+\; C \frac{h^{2\mu-3}_{K_2}}{r^{2s-3}} \|\psi\|^2_{s,K_2} \\
&\quad +\; C\, h^{-1}_{K_1} r \, \|\nabla\psi - \nabla\psi_h\|^2_{0,K_1} \;+\; C\, h^{-1}_{K_2} r \, \|\nabla\psi - \nabla\psi_h\|^2_{0,K_2} \\
&\leq\; C \frac{h^{2\mu-3}}{r^{2s-3}} \|\psi\|^2_{s,K_{12}} \;+\; C\, h^{-1} r \, \|\nabla\psi - \nabla\psi_h\|^2_{0,K_{12}} \qquad (3.68)
\end{aligned}$$

The above inequality and (3.62) together imply that

$$\begin{aligned}
\| [\nabla\psi - \nabla\psi_h] \|^2_{0,\mathring{\mathcal{F}}_h} \;&=\; \sum_{k=1}^{P_h} \| [\nabla\psi - \nabla\psi_h] \|^2_{0,f_k} \\
&\leq\; C \frac{h^{2\mu-3}}{r^{2s-3}} \sum_{k=1}^{P_h} \|\psi\|^2_{s,K_{12}} \;+\; C\, h^{-1} r \sum_{k=1}^{P_h} \|\nabla\psi - \nabla\psi_h\|^2_{0,K_{12}} \\
&\leq\; C \frac{h^{2\mu-3}}{r^{2s-3}} \| \psi \|^2_{s,\Omega} \;+\; C\, h^{-1} r \, \|\nabla\psi - \nabla\psi_h\|^2_{0,\Omega} \,. \\
&\leq\; C \frac{h^{2\mu-3}}{r^{2s-3}} \| \psi \|^2_{s,\Omega} \,. \qquad (3.69)
\end{aligned}$$

The show (3.66), let $f_k = K_1 \cap K_2$ be an arbitray interior face. Then

$$\begin{aligned}
\| \{\nabla\psi - \nabla\psi_h\} \|^2_{0,f_k} \;&\leq\; \big(\frac{1}{2} \| (\nabla\psi - \Pi^r_h \psi)_{|K_1} \|_{0,f_k} \;+\; \frac{1}{2} \| (\nabla\psi - \Pi^r_h \psi)_{|K_2} \|_{0,f_k} \\
&\quad +\; \frac{1}{2} \| (\nabla\psi_h - \Pi^r_h \psi)_{|K_1} \|_{0,f_k} \;+\; \frac{1}{2} \| (\nabla\psi_h - \Pi^r_h \psi)_{|K_2} \|_{0,f_k} \big)^2 . \qquad (3.70)
\end{aligned}$$

At this point, the same argument that was used to prove (3.65) can be applied to the above inequality in order to show that (3.66) holds.

To show that the estimates (3.65)-(3.66) are optimal in both h and r, we note that from

interpolation properties (2.96) it readily follows that the interpolant $\Pi_h^r \psi$ satisfies

$$\| [\nabla\psi - \nabla(\Pi_h^r \psi)] \|_{0,\mathring{\mathcal{F}}_h} \leq C \frac{h^{\mu-\frac{3}{2}}}{r^{s-\frac{3}{2}}} \| \psi \|_{s,\Omega} \qquad \text{and} \tag{3.71}$$

$$\| \{\nabla\psi - \nabla(\Pi_h^r \psi)\} \|_{0,\mathring{\mathcal{F}}_h} \leq C \frac{h^{\mu-\frac{3}{2}}}{r^{s-\frac{3}{2}}} \| \psi \|_{s,\Omega} . \tag{3.72}$$

Thus, the estimates (3.65) and (3.66) satisfied by ψ_h attain the same hp-convergence order as the above estimates (3.71)-(3.72) satisfied by $\Pi_h^r \psi$. Hence, they are optimal in both h and r. □

3.4 *A priori* NIPG error estimate for the perturbed Poisson system

We now consider the NIPG approximation of the Poisson system, for a fixed mesh $\mathcal{T}_h$, that is defined by a righthandside source term $\beth_h \in L^2(\Omega)$, which is indexed by h since this function is allowed to be mesh dependent. In particular, we will investigate how the error between the discrete NIPG solution ψ_h for this system and the true NIPG solution ψ for the Poisson system defined by a source term $F \in L^2(\Omega)$, i.e., $\| \psi - \psi_h \|_{NIPG}$, depends on $\| F - \beth_h \|_{L^2(\Omega)}$.

In the rest of this subsection, we assume that $\Gamma_D = \partial\Omega$, and hence $\Gamma_N = \emptyset$. For a given source term $\beth_h$, the corresponding Poisson system defined by this source term is then stated as:

$$-\nabla \cdot (A \nabla \tilde{\psi}_h) = \beth_h \qquad \text{in } \Omega , \tag{3.73}$$

$$\tilde{\psi}_h = r_D \qquad \text{on } \Gamma_D . \tag{3.74}$$

The above system will be referred to as the perturbed Poisson system. The domain, boundary data, and matrix A for the perturbed system are exactly the same quantities as those used in (3.1)-(3.3) for the Poisson system defined by the source function F.

By Theorem 10, it follows that there exists unique distributional solutions $\psi = \psi(F) \in H^1(\Omega)$ and $\tilde{\psi}_h = \tilde{\psi}_h(\beth_h) \in H^1(\Omega)$ to the regular and the perturbed Poisson systems, respectively. Moreover, throughout the rest of this section, it will be assumed that ψ and $\tilde{\psi}_h$ both satisfy the additional regularity condition that they are both contained in the function space $H^1(\Omega) \cap H^2(\mathcal{T}_h)$. Under this assumption, ψ and $\tilde{\psi}_h$ are unique, true NIPG solutions

to the regular and perturbed Poisson systems, respectively. We remark that from the above assumptions, if follows that

$$\psi \; - \; \tilde{\psi}_h \; = \; 0 \, , \qquad \text{on } \Gamma_{_D} \, . \tag{3.75}$$

It will be seen that this result is crucial in the analysis to come.

We now state the following theorem concerning the error between the Poisson system and the discrete perturbed Poisson system.

Theorem 13. *Suppose $\Gamma_{_D} = \partial\Omega$. Let ψ be the distributional solution to the Poisson system defined by a source function $F \in L^2(\Omega)$, let $\tilde{\psi}_h$ be the distributional solution to the perturbed Poisson system defined by a source function $\mathfrak{I}_h \in L^2(\Omega)$,, and let ψ_h be the discrete NIPG approximation to $\tilde{\psi}_h$. If ψ, $\tilde{\psi}_h \in H^1(\Omega) \cap H^s(\mathcal{T}_h)$, for $s \geq 2$, and $\beta = \frac{1}{2}$, then the following is satisfied:*

$$\begin{aligned} &\| A^{1/2} \nabla(\psi - \psi_h) \|^2_{0,\Omega} \; + \; \sum_{k=1}^{P_h} \frac{r\sigma_k}{| f_k |^{\frac{1}{2}}} \| [\psi - \psi_h] \|^2_{0,f_k} \; + \; \sum_{f_k \in \Gamma_{_D}} \frac{r\sigma_k}{| f_k |^{\frac{1}{2}}} \| \psi - \psi_h \|^2_{0,f_k} \\ &\leq \; \lambda^{-1}_{min} \, \| F - \mathfrak{I}_h \|^2_{L^2(\Omega)} \; + \; C \, \frac{h^{2\mu-2}}{r^{2s-2}} \| \tilde{\psi}_h \|^2_{s,\Omega} \, , \end{aligned} \tag{3.76}$$

where λ_{min} is the constant given in (3.31).

Proof. The procedure for estimating $\| \psi - \psi_h \|_{NIPG}$ is to split this estimate up into two parts as follows:

$$\| \psi - \psi_h \|^2_{NIPG} \; \leq \; 2 \| \psi - \tilde{\psi}_h \|^2_{NIPG} \; + \; 2 \| \tilde{\psi}_h - \psi_h \|^2_{NIPG} . \tag{3.77}$$

We now handle each of the above terms seperately.

Since ψ_h is the discrete NIPG solution of the function $\tilde{\psi}_h$, we get from Theorem 12 that the following estimate is satisfied:

$$\| \tilde{\psi}_h - \psi_h \|_{NIPG} \; \leq \; C \, \frac{h^{2\mu-2}}{r^{2s-2}} \| \tilde{\psi}_h \|^2_{s,\Omega} \, . \tag{3.78}$$

What now remains is to establish control over the quantity $\| \psi - \tilde{\psi}_h \|_{NIPG}$. To do this, we first note since ψ is a true NIPG solution to the Poisson system defined by F, we get that

$$a(\, \psi, \, w \,) \; + \; J(\, \psi, \, w \,) \; = \; L(\, w; \, F \,)$$

is satisfied $\forall\ w \in H^1(\mathcal{T}_h)$, where the functional notation $L(\,\cdot;\,F\,)$ expresses the dependence of L on the function F. Similarly, for $\tilde{\psi}_h$ we have that

$$a(\,\tilde{\psi}_h,\,w\,)\ +\ J(\,\tilde{\psi}_h,\,w\,)\ =\ L(\,w;\,\mathfrak{I}_h\,)$$

is satisfied $\forall\ w \in H^1(\mathcal{T}_h)$. Upon setting the test function $w = \psi - \tilde{\psi}_h$ and then differencing the above equations leads to the inequality

$$\begin{aligned}
\|\,\psi - \tilde{\psi}_h\,\|^2_{NIPG} &= \int_\Omega (\,F - \mathfrak{I}_h\,)\,(\,\psi - \tilde{\psi}_h\,)\,dx \\
&\leq \frac{\epsilon}{2}\,\|\,F - \mathfrak{I}_h\,\|^2_{L^2(\Omega)}\ +\ \frac{1}{2\epsilon}\,\|\,\psi - \tilde{\psi}_h\,\|^2_{L^2(\Omega)} \\
&\leq \frac{\epsilon}{2}\,\|\,F - \mathfrak{I}_h\,\|^2_{L^2(\Omega)}\ +\ \frac{1}{2\epsilon}\,\|\,\nabla(\psi - \tilde{\psi}_h)\,\|^2_{L^2(\Omega)}\,, \qquad (3.79)
\end{aligned}$$

for any $\epsilon > 0$. The last line of the above inequality follows from Poincare's inequality, which we can take advantage of since $\psi - \tilde{\psi}_h \in H^1_0(\Omega)$.

Now recall the property (3.31) of the matrix A. This property implies that

$$\|\,\xi\,\|^2_{L^2(\Omega)}\ \leq\ \lambda^{-1}_{min}\,\|A^{1/2}\,\xi\|^2_{L^2(\Omega)}\,, \qquad \forall\ \xi \in \mathbb{R}^3\,.$$

Using this result, we then get that

$$\begin{aligned}
\|\,\psi - \tilde{\psi}_h\,\|^2_{NIPG} &\leq \frac{\epsilon}{2}\,\|\,F - \mathfrak{I}_h\,\|^2_{L^2(\Omega)}\ +\ \frac{\lambda^{-1}_{min}}{2\epsilon}\,\|\,A^{1/2}\,\nabla(\psi - \tilde{\psi}_h)\,\|^2_{L^2(\Omega)} \\
&\leq \frac{1}{2\lambda_{min}}\,\|\,F - \mathfrak{I}_h\,\|^2_{L^2(\Omega)}\ +\ \frac{1}{2}\,\|\,A^{1/2}\,\nabla(\psi - \tilde{\psi}_h)\,\|^2_{L^2(\Omega)}\,, \qquad (3.80)
\end{aligned}$$

where ϵ is chosen to be to the λ^{-1}_{min}.

By definition, we know that

$$\begin{aligned}
\|\,\psi - \tilde{\psi}_h\,\|^2_{NIPG} &= \|\,A^{1/2}\,\nabla(\psi - \tilde{\psi}_h)\,\|^2_{L^2(\Omega)}\ +\ J(\,\psi - \tilde{\psi}_h,\,\psi - \tilde{\psi}_h\,) \\
&= \|\,A^{1/2}\,\nabla(\psi - \tilde{\psi}_h)\,\|^2_{L^2(\Omega)}\,, \qquad (3.81)
\end{aligned}$$

where the penalty term vanishes due to the regularity of ψ and $\tilde{\psi}_h$ and the fact that $\psi = \tilde{\psi}_h$ on $|\Gamma_D|$. After combining the above inequality with the previous inequality, it follows

$$\|\,\psi - \tilde{\psi}_h\,\|^2_{NIPG}\ \leq\ \frac{1}{\lambda_{min}}\,\|\,F - \mathfrak{I}_h\,\|^2_{L^2(\Omega)}\,. \qquad (3.82)$$

Upon combining inequalties (3.78) and (3.82),

$$\| \psi - \psi_h \|^2_{NIPG} \;\leq\; \lambda^{-1}_{min} \| F - \beth_h \|^2_{L^2(\Omega)} \;+\; C \, \frac{h^{2\mu-2}}{r^{2s-2}} \| \tilde{\psi}_h \|^2_{L^2(\Omega)} \, . \tag{3.83}$$

□

Chapter 4

DG methods for the Vlasov and Vlasov-Poisson systems

4.1 Introduction

Designing a discretization method that gives a constructive way of finding a finite dimensional function that approximates an uknown solution to the Vlasov-Poisson system whose phase-space is a bounded subspace of $\mathbb{R}^3 \times \mathbb{R}^3$ is a daunting task. This stems from the fact that the problem is a highly nonlinear problem that evolves in time over a six-dimensional space. Moreover, if it is desired that the method be consistent and amenable to establishing rigorous stability and error estimates, then the task of designing the method becomes even more burdersome. The goal of this chapter is to propose a DG approximation to the Vlasov-Poisson system that is consistent and is such that an *a priori* error estimate can be derived.

The goals set forth will be achieved by first considering the problem of finding a DG discretization of the Vlasov system, where the potential, and hence the electric field, are assumed to be given. With these assumptions, our problem resembles a standard transport problem, except for the excessive dimensionality of the Vlasov system and the specific structure of its the flow vector. Hence, our starting point is to consider the well-known upwind Galerkin (UG) method of approximation, which is a DG method for the discretization of conservation laws [34], [33], [30], [31], [35], [29], [4].

In anticipation of the fact that the electric field will be approximated by a discrete function that is discontinuous across the mesh faces, we will introduce a new DG method to discretize

the Vlasov system, in the case when the potential, and hence the flow field defining the Vlasov equation, are discontinuous across the mesh faces. The starting point for developing this method is to investigate the UG method, so that we can understand how it can be modified to allow for flows that are discontinuous. The method that results from modifying the UG method will be called the *discontinuous flow upwind Galerkin* (DFUG) method. The DFUG method will be shown to be stable and consistent.

After introducing the DFUG method, we will consider a Vlasov system that is defined by a flow α and a perturbed Vlasov system that is defined by a flow $\aleph_h$, which is some perturbation of α. We then investigate the error between the solutions to the DFUG formulation of the regular Vlasov system and the perturbed Vlasov system. An explicit *a priori* error estimate will be derived for the difference between these two solutions. This estimate will be seen to contain contributions coming from both the DFUG discretization of the Vlasov equation and the error coming from the difference between the regular and perturbed flow vectors. In the case when the perturbed flow α is set equal to $\aleph_h$, the derived error estimate will be seen to be optimal in h.

After deriving an error estimate for the DFUG method, in which the final estimate depends explicitly on normed quantities of the difference between the regular flow and the perturbed flow, we then assume that the perturbed flow field satisfies a typical hp-error estimate result. We then plug this result into the error estimate obtained for the DFUG method, which results in a new error estimate. In particular, this result will show what order of error in h needs to be obtained by the approximation error between the regular and the perturbed flow fields in order to maintain an error estimate that is optimal in h.

The final work in this section will involve combining the results from Chapter 3 for the NIPG approximation to the Poisson system and the results for the DFUG approximation to the Vlasov system in such a way that an error estimate can be derived for the DG method proposed to approximate the Vlasov-Poisson system. This proposed method is called the discontinuouos flow upwind Galerkin - nonsymmetric interior penalty Galerkin (DFUG-NIPG) method, since it combines the two respective methods together in order to give a nonlinear, phase space, DG discretization of the Vlasov-Poisson system.

4.2 Mesh structure

The notation used to describe the mesh structure introduced in Chapter 2 and utilized in Chapter 3 assumed that the underlying domain Ω was a bounded subset in $\mathbb{R}^3$. In this

chapter, new notation is introduced in order to handle the fact that $\Omega \subset \mathbb{R}^6$. However, all of the definitions and mesh properties that were assumed in Chapter 2 are still valid, but need to be updated to reflect the higher dimensionality and the nature of Ω for the Vlasov-Poisson problem.

In all of the work to come, it will be assumed that the domain Ω can be partitioned as

$$\Omega \;=\; \Omega^{x} \cup \Omega^{v} \;,$$

where

$$\Omega^{x} \;=\; [\,0,\, L_1\,] \times [\,0,\, L_2\,] \times [\,0,\, L_3\,] \quad \text{and} \quad \Omega^{v} \;=\; [\,-V,\, V\,]^3 \;,$$

where $L_1, L_2, L_3, V > 0$ are fixed constants. The role of the constant V is that it will always be assumed that any true solution to the Vlasov-Poisson system being considered satisfies the condition

$$\sup\{\,|v| \;:\; f(\,x, v, t\,) \neq 0\,,\; x \in \Omega^{x},\; 0 \leq t \leq T\,\} \;<\; V\,,$$

i.e., the velocity support of f is contained within the domain Ω^{v}.

Let $\{\,\mathcal{T}_h\,\}_{h>0}$ be a family of meshes for the domain Ω, where for each mesh $\mathcal{T}_h = \{\,K_1, \ldots, K_{N_h}\,\}$ it is assumed that

$$(i) \quad \overline{\Omega} \;=\; \cup_{j=1}^{N_h}\, K_j\,, \tag{4.1}$$

$$(ii) \quad K_j \text{ is compact, convex, for } j = 1, \ldots, N_h\,, \quad \text{and} \tag{4.2}$$

$$(iii) \quad \mathring{K}_i \cap \mathring{K}_j \;=\; \emptyset\,, \quad \text{for } i \neq j,\; i, j = 1, \ldots, N_h\,. \tag{4.3}$$

In order to maintain tractibility during the DG analysis for the Vlasov-Poisson problem, we must tailor the mesh so that it incorporates the structure of Ω.

To begin with, we assume there exists a family of meshes $\{\,\mathcal{T}_h^{x}\,\}_{h>0}$ for the domain Ω^{x} and a family of meshes $\{\,\mathcal{T}_h^{v}\,\}_{h>0}$ for the domain Ω^{v} such that for each $\mathcal{T}_h$ we have that

$$\mathcal{T}_h \;=\; \mathcal{T}_h^{x} \times \mathcal{T}_h^{v}\,. \tag{4.4}$$

Each of meshes $\mathcal{T}_h^{x} = \{\,K_1^{x}, \ldots, K_{N_h^{x}}^{x}\,\}$ and $\mathcal{T}_h^{v} = \{\,K_1^{v}, \ldots, K_{N_h^{v}}^{v}\,\}$, is assumed to satisfy

analogous properties to (4.1)-(4.3). Clearly, the above assumptions imply that

$$\mathcal{T}_h \;=\; \{\, K^x_{j_x} \times K^v_{j_v} \,\}^{N^x_h, N^v_h}_{j_x, j_v = 1} \,, \tag{4.5}$$

so that $N_h = N^x_h \, N^v_h$.

For each element $K_j \in \mathcal{T}_h$, we define its diameter in the usual manner $h_j \;=\; \mathrm{diam}(\, K_j \,)$. We then define the maximum diameter h to be maximum diameter among all of the elements. For the meshes $\mathcal{T}^x_h$ and $\mathcal{T}^v_h$, where $\mathcal{T}_h \;=\; \mathcal{T}^x_h \cup \mathcal{T}^v_h$, we define the quantities, for a given element $K_j = K^x_{j_x} \cup K^v_{j_v} \in \mathcal{T}_h$,

$$h_{j_x, x} \;=\; \mathrm{diam}(\, K^x_{j_x} \,) \qquad \text{and} \qquad h_{j_v, v} \;=\; \mathrm{diam}(\, K^v_{j_v} \,) \;. \tag{4.6}$$

We then define h_x and h_v to be the maximum diameters among all of the elements in $\mathcal{T}^x_h$ and $\mathcal{T}^v_h$, respectively. With these definitions, it then follows that

$$\begin{aligned} h_j \;&=\; \max_{(x_1, v_1), (x_2, v_2) \in K_j} |\, (x_1, v_1) - (x_2, v_2) \,|_{\mathbb{R}^6} \\ &=\; \max_{x_1, x_2 \in K^x_{j_x}, \; v_1, v_2 \in K^v_{j_v}} \sqrt{|\, x_1 - x_2 \,|^2_{\mathbb{R}^3} \;+\; |\, v_1 - v_2 \,|^2_{\mathbb{R}^3}} \\ &=\; \sqrt{h^2_{j_x, x} \;+\; h^2_{j_v, v}} \;. \end{aligned} \tag{4.7}$$

We will see later on that the above relationship can be further simplified.

That $\mathcal{T}_h$ is an affine is say that every element K_j is the image of the unit cube

$$\widehat{K} \;=\; [\, 0, 1 \,]^6 \tag{4.8}$$

under an affine, bijective transformation T_{K_j}. Moreover, T_{K_j} is defined by two affine, bijective transformations $T^x_{K^x_{j_x}}$ and $T^v_{K^v_{j_v}}$, such that there exist constant invertible matrices $M^x_{K^x_{j_x}}$, $M^v_{K^v_{j_v}} \in \mathbb{R}^{3\times 3}$ and constant vectors $b^x_{K^x_{j_x}}$, $b^v_{K^v_{j_v}} \in \mathbb{R}^{3\times 3}$ satisfying

$$T_{K_j} : \widehat{K} \ni (\, \hat{x}, \hat{v} \,) \;\longrightarrow\; (\, x, v \,) \;=\; (\, T^x_{K^x_{j_x}} \, \hat{x}, \; T^v_{K^v_{j_v}} \, \hat{v} \,) \in K_j \;, \tag{4.9}$$

where $T^x_{K^x_{j_x}} \, \hat{x} \;=\; M^x_{K^x_{j_x}} \, \hat{x} \;+\; b^x_{K^x_{j_x}}$ and $T^v_{K^v_{j_v}} \, \hat{v} \;=\; M^v_{K^x_{j_x}} \, \hat{v} \;+\; b^v_{K^v_{j_v}}$.

The notion of a mesh face gets considerably more complicated as the dimensionality in-

creases. This seen is from the fact the reference element $[0,1]^6$ contains twelve faces:

$$\begin{aligned}
\hat{f}_1 &= \{0, x_2, x_3, v_1, v_2, v_3\} \quad , \quad \hat{f}_2 = \{1, x_2, x_3, v_1, v_2, v_3\} \\
\hat{f}_3 &= \{x_1, 0, x_3, v_1, v_2, v_3\} \quad , \quad \hat{f}_4 = \{x_1, 1, x_3, v_1, v_2, v_3\} \\
\hat{f}_5 &= \{x_1, x_2, 0, v_1, v_2, v_3\} \quad , \quad \hat{f}_6 = \{x_1, x_2, 1, v_1, v_2, v_3\} \\
\hat{f}_7 &= \{x_1, x_2, x_3, 0, v_2, v_3\} \quad , \quad \hat{f}_8 = \{x_1, x_2, x_3, 1, v_2, v_3\} \\
\hat{f}_9 &= \{x_1, x_2, x_3, v_1, 0, v_3\} \quad , \quad \hat{f}_{10} = \{x_1, x_2, x_3, v_1, 1, v_3\} \\
\hat{f}_{11} &= \{x_1, x_2, x_3, v_1, v_2, 0\} \quad , \quad \hat{f}_{12} = \{x_1, x_2, x_3, v_1, v_2, 1\} \, ,
\end{aligned} \tag{4.10}$$

where $0 \leq x_1, x_2, x_3, v_1, v_2, v_3 \leq 1$. So, the faces of any particular element K_j are given by $T_{K_j}(\hat{f}_k)$, $k = 1, \ldots, 12$. As in Chapter 2, the mesh faces f_k are defined to be the non-empty intersections of the faces of adjacent elements. The set of all mesh faces is denoted by

$$\mathcal{F}_h = \{ f_1, \ldots, f_{P_h}, \ldots, f_{M_h} \} , \tag{4.11}$$

where f_k is an interior face for $k = 1, \ldots, P_h$, and f_k is a boundary face for $k = P_h + 1, \ldots, M_h$. The set of all interior faces $\mathring{\mathcal{F}}_h$ is then given by

$$\mathring{\mathcal{F}}_h = \{ f_1, \ldots, f_{P_h} \} . \tag{4.12}$$

The fact that we assume for every element $K_j \in \mathcal{T}_h$ there exist $K^x_{j_x} \in \mathcal{T}^x_h$ and $K^v_{j_v} \in \mathcal{T}^v_h$ such that $K_j = K^x_{j_x} \times K^v_{j_v}$ implies that $\mathcal{F}_h$ can be decomposed in a natural way, which we now discuss.

Let us define the set of mesh faces for $\mathcal{T}^x_h$ and $\mathcal{T}^v_h$, respectively, as was done in Chapter 2 for bounded domains that are subsets of $\mathbb{R}^3$. We denote the set of faces for $\mathcal{T}^x_h$ by

$$\mathcal{G}^x_h = \{ f^x_1, \ldots, f^x_{P^x_h}, \ldots, f^x_{M^x_h} \} \tag{4.13}$$

and we denote the set of faces for $\mathcal{T}^v_h$ by

$$\mathcal{G}^v_h = \{ f^v_1, \ldots, f^v_{P^v_h}, \ldots, f^v_{M^v_h} \} , \tag{4.14}$$

where $f^x_{k_x}$ is an interior face of $\mathcal{G}^x_h$ for $k_x = 1, \ldots, P^x_h$ and is a boundary face for $k_x = P^x_h + 1, \ldots, M^x_h$ and $f^v_{k_v}$ is an interior face of $\mathcal{G}^v_h$ for $k_v = 1, \ldots, P^v_h$ and is a boundary face

for $k_v = P_h^v + 1, \ldots, M_h^v$. We define the following the interior face sets

$$\mathring{\mathcal{G}}_h^x = \{ f_1^x, \ldots, f_{P_h^x}^x \} \qquad \text{and} \tag{4.15}$$

$$\mathring{\mathcal{G}}_h^v = \{ f_1^v, \ldots, f_{P_h^v}^v \}. \tag{4.16}$$

For every $f_{k_x}^x \in \mathcal{G}_h^x$ and $f_{k_v}^v \in \mathcal{G}_h^v$, we associate the unit normal vectors $\nu_{k_x}^x$ and $\nu_{k_v}^v$, respectively. For $k_x > P_h^x$, $\nu_{k_x}^x$ is taken to be the outward unit normal vector to $\partial\Omega^x$. For $k_v > P_h^v$, $\nu_{k_v}^v$ is taken to be the outward unit normal vector to $\partial\Omega^v$. We note for future reference that $\nu_{K_{j_x}^x}^v$ will be used to denote the outward unit normal vector to $\partial K_{j_x}^x$ and $\nu_{K_{j_v}^v}^v$ will be used to denote the outward unit normal vector to $\partial K_{j_v}^v$. Also, for interior faces $f_{k_x}^x = K_1^x \cap K_2^x$ and$f_{k_v}^v = K_1^v \cap K_2^v$ it will be assumed that K_1^x and K_1^v are the elements such that $\nu_{K_1^x}^x = \nu_{k_x}^x$ on $f_{k_x}^x$ and $\nu_{K_1^v}^v = \nu_{k_v}^v$ on $f_{k_v}^v$, which then implies $\nu_{K_2^x}^x = -\nu_{k_x}^x$ on $f_{k_x}^x$ and $\nu_{K_2^v}^v = -\nu_{k_v}^v$ on $f_{k_v}^v$.

A face $f_k \in \mathcal{F}_h$ is such that either (i) there exists $f_{k_x}^x \in \mathcal{G}_h^x$ and $K_{j_v}^v \in \mathcal{T}_h^v$ such that $f_k = f_{k_x}^x \times K_{j_v}^v$ or (ii) there exists $K_{j_x}^x \in \mathcal{T}_h^x$ and $f_{k_v}^v \in \mathcal{G}_h^v$ such that $f_k = K_{j_x}^x \times f_{k_v}^v$. We define the sets $\mathcal{F}_h^x$ and $\mathcal{F}_h^v$ by

$$\begin{aligned}
\mathcal{F}_h^x &= \{ f_k \in \mathcal{F}_h : \ f_k = f_{k_x}^x \times K_{j_v}^v, \text{ for some } \ f_{k_x}^x \in \mathcal{G}_h^x \text{ and } K_{j_v}^v \in \mathcal{T}_h^v \} \qquad \text{and} \\
\mathcal{F}_h^v &= \{ f_k \in \mathcal{F}_h : \ f_k = K_{j_x}^x \times f_{k_v}^v, \text{ for some } \ K_{j_x}^x \in \mathcal{T}_h^x \text{ and } f_{k_v}^v \in \mathcal{G}_h^v \}.
\end{aligned}$$

We then define the interiors of these sets by

$$\begin{aligned}
\mathring{\mathcal{F}}_h^x &= \{ f_k \in \mathcal{F}_h : \ f_k = f_{k_x}^x \times K_{j_v}^v, \text{ for some } \ f_{k_x}^x \in \mathring{\mathcal{G}}_h^x \text{ and } K_{j_v}^v \in \mathcal{T}_h^v \} \qquad \text{and} \\
\mathring{\mathcal{F}}_h^v &= \{ f_k \in \mathcal{F}_h : \ f_k = K_{j_x}^x \times f_{k_v}^v, \text{ for some } \ K_{j_x}^x \in \mathcal{T}_h^x \text{ and } f_{k_v}^v \in \mathring{\mathcal{G}}_h^v \}.
\end{aligned}$$

It follows from these definitions that we have the partitions

$$\mathcal{F}_h = \mathcal{F}_h^x \cup \mathcal{F}_h^v$$

and

$$\mathring{\mathcal{F}}_h = \mathring{\mathcal{F}}_h^x \cup \mathring{\mathcal{F}}_h^v.$$

These partitions will be utilized during the forthcoming DG analysis for the Vlasov-Poisson system.

The mesh properties of regularity and quasi-uniformity are the same as originally stated in

Lemmas 1 and 2, with some minor modifications. In Lemma 1, (2.40) still holds, where $\rho_{\widehat{K}}$ is now the largest ball that can be inscribed in $\widehat{K} = [0,1]^6$. An analogous result to Lemma 2 can be derived in the same manner and is stated as follows:

Lemma 15. *Let $K_j \in \mathcal{T}_h$ be arbitrary. Let f_{K_j} be an arbitrary face of K_j and let e_{K_j} be an arbitrary edge of K_j. Then*

$$|K_j| \sim h_j^6 \sim h^6, \qquad |f_{K_j}| \sim h_j^5 \sim h^5, \qquad |e_{K_j}| \sim h_j^4 \sim h^4. \tag{4.17}$$

One last property of the mesh to note is that the regularity condition implies that, $\forall\, \mathcal{T}_h = \mathcal{T}_h^x \times \mathcal{T}_h^v$,

$$h_j \sim h_{j_x,x} \sim h_{j_v,v}, \quad \forall\, K_j \in \mathcal{T}_h, \qquad \text{and hence} \tag{4.18}$$

$$h \sim h_x \sim h_v. \tag{4.19}$$

The above equivalencies imply that there exist a mesh-independent constants $M_{xv} > 0$ and $M_{vx} > 0$ satisfying, $\forall\, h > 0$, $h_v \leq M_{xv}\, h_x$ and $h_x \leq M_{vx}\, h_v$. This condition prevents the situation where, as $h \to 0$, one mesh, say $\mathcal{T}_h^v$, is refined indefinitely, while the other mesh $\mathcal{T}_h^x$ is not refined at all. Stated another way, there is an upper bound and lower bound on how much the refinement levels of $\mathcal{T}_h^x$ and $\mathcal{T}_h^v$ can differ.

4.3 Vlasov and Vlasov-Poisson systems of equations

In this chapter, we consider the Vlasov-Poisson system of equations, which is a nonlinear system that can be used to model the transport of both electrons and ions within a collisionless, or near collisionless, plasma. Our goal is to propose a DG method for the approximation of this system. Moreover, we want the proposed method to be consistent, meaning that if a classical solution exists then it also satisfies the weak formulation defining the DG method, and to be such that *a priori* error estimates can be proved.

The Vlasov-Poisson system of equations that will be considered in this chapter will be now clearly defined. In order to maintain simplicity, we do not include an external forcing function in the Poisson equation and we assume that the diffusion matrix A of the Poisson equation is equal to I, the identity matrix. However, all of the results concerning this

particular Vlasov-Poisson system hold if an external forcing function were present and if A was not the identity, but instead satisfied the matrix properties outlined in Chapter 3 for the Poisson equation.

Definition 9. [Vlasov – Poisson System of Equations] *Given $T > 0$ and a data trio (f_0, f_I, r_D), the Vlasov system of equations up to time T is defined to be the set of equations*

$$\begin{aligned} f_t \; + \; \alpha \cdot \nabla f \;\; &= \;\; 0 \,, && \Omega \times (0,T] \,, && (4.20) \\ E \;\; &= \;\; -\nabla_x \psi \,, && \Omega^x \times (0,T] \,, && (4.21) \\ -\Delta_x \psi \;\; &= \;\; \rho(f) \,, && \Omega^x \times (0,T] \,, && (4.22) \\ f(\, t = 0 \,) \;\; &= \;\; f_0 \,, && \Omega \,, && (4.23) \\ f \;\; &= \;\; f_I \,, && \Gamma_I(t) \times (0,T] \,, && (4.24) \\ \psi \;\; &= \;\; r_D \,, && \partial\Omega^x \times (0,T] \,, && (4.25) \end{aligned}$$

where

$$\Omega \;\; = \;\; \Omega^x \times \Omega^v \;\; = \;\; [\, 0,\, L_1 \,] \times [\, 0,\, L_2 \,] \times [\, 0,\, L_3 \,] \times [\, -V,\, V \,]^3 \,, \tag{4.26}$$

with L_1, L_2, L_3, V being fixed positive constants, and where

$$\alpha(x,v,t) \;\; = \;\; \begin{pmatrix} v \\ \nabla_x \psi(x,t) \end{pmatrix} . \tag{4.27}$$

As mentioned earlier, in order to truncate the velocity domain as we have done requires that we assume that the initial condition f_0 has compact velocity support in $[\, -V,\, V \,]^3$ and the true solution f, assuming it exists, has compact velocity support in $[\, -V,\, V \,]^3$, for all times $t \in (0,T]$. With the velocity domain now bounded, the inflow boundary is defined as

$$\Gamma_I(t) \;\; = \;\; \{\, (x,v) \in \partial\Omega : \;\; \alpha(x,v,t) \cdot \nu \;\; < \;\; 0 \} \,, \tag{4.28}$$

where

$$\partial\Omega \;\; = \;\; (\, \partial\Omega^x \times \Omega^v \,) \times (\, \Omega^x \times \partial\Omega^v \,) \,.$$

We note that the inflow boundary is time dependent when the velocity domain is bounded, whereas it is time independent when $\Omega^v = \mathbb{R}^3$.

The corresponding data compatibility conditions for the above system of equations (4.20)-

(4.25) are defined as follows:

Definition 10. [Data Compatibility for Vlasov – Poisson System] *Given $T > 0$, the data trio (f_0, f_I, r_D) is said to be compatible up to time T if*

$$\begin{aligned} f_0 &\in C^1(\,\Omega^x,\, C_c^1(\Omega^v)\,)\,, && (4.29)\\ f_I &\in C^1(\,\partial\Omega \times [0,T]\,)\,, && (4.30)\\ f_I(\cdot, v, \cdot) &= 0\,, \quad \forall\ v \in \partial\Omega^v\,, && (4.31)\\ \partial^{|\,\beta\,|} f_I(t=0) &= \partial^{|\,\beta\,|} f_0\,, \quad \text{on } \Gamma_I(t=0),\ \forall\ |\,\beta\,| \ \leq\ 1\,, \text{ and} && (4.32)\\ r_D &\in L^2(\partial\Omega^v)\,, && (4.33) \end{aligned}$$

where β is a multi-index.

During the DG error analysis of the Vlasov-Poisson system, it will always be assumed that the given data trio for defining the system is compatible according to Definition 10 and that there exists a unique classical solution satisfying equations (4.20)-(4.25). Moreover, regularity properties of the solution f will also be assumed, i.e., $f \in C^1(\,[0,T],\, C^4(\Omega)\,)$. Strong regularity assumptions on f are reasonable provided that the given problem data are smooth enough. This follows from the fact the solutions f and ψ of the Vlasov-Poisson system inherit their regularity from the data defining the system.

In order to investigate the DG methods of approximation to the Vlasov-Poisson system described above, we will first investigate the simpler Vlasov system of equations. This type of approach to analyzing the Vlasov-Poisson system, that of considering the Vlasov and Poisson systems seperately at first, follows the approaches taken in [61], [66], and [53], as outlined in Chapter 2, to prove regularity results. Each of the two seperate systems are linear, when they are considered independently of each other, so that many results are known or can be shown, if need be, for each of these systems. The trick to combining the linear results together to get results for the nonlinear system is to then use an iterated sequence of solution pairs (f^n, ψ^n), as discussed in Chapter 2, whose convergence to a unique solution pair can be established.

Definition 11. [Vlasov System of Equations] *Given $T > 0$ and a data trio $(\,(\alpha, f_0, f_I)\,)$, the Vlasov system of equations up to time T is defined to be the set of equations*

$$\begin{aligned} f_t \ +\ \alpha \cdot \nabla f &= 0\,, && \Omega \times (0,T]\,, && (4.34)\\ f(\,t=0\,) &= f_0\,, && \Omega\,, && (4.35)\\ f &= f_I\,, && \Gamma_I(t) \times (0,T]\,, && (4.36) \end{aligned}$$

where α, Ω, *and* $\Gamma_I(t)$ *are as specified above.*

We note that the Vlasov equation (4.34) is identical to a standard linear tranport equation, except that the flow vector α has the additional structure coming from its definition given in (4.27). This extra structure of α will be exploited in the remainder of this work, thereby making the forthcoming analysis not directly applicable to general linear transport equations. However, an attempt is made to be as general as possible in the presentation of the material to come, so that the ideas and concepts here can be easily adapted to more general transport equations.

We now state the compatibility conditions that will be assumed to hold for the flow and the initial and boundary data that define the Vlasov system.

Definition 12. [Data Compatibility for Vlasov System] *Given* $T > 0$, *the data trio* (α, f_0, f_I) *is said to be compatible up to time* T *if*

$$
\begin{aligned}
f_0 &\in C^1(\,\Omega^x,\, C_c^1(\Omega^v)\,)\,, && (4.37)\\
f_I &\in C^1(\,\partial\Omega \times [0,T]\,)\,, && (4.38)\\
f_I(\cdot, v, \cdot) &= 0\,, \quad \forall\ v \in \partial\Omega^v\,, && (4.39)\\
\partial^{|\,\beta\,|} f_I(t=0) &= \partial^{|\,\beta\,|} f_0\,, \quad \text{on } \Gamma_I(t=0),\ \forall\ |\,\beta\,| \ \le\ 1\,, \text{ and} && (4.40)\\
\psi &\in C^1(\,[0,T],\, C^2(\Omega^x)\,)\,. && (4.41)
\end{aligned}
$$

The compatibility conditions for both the Vlasov and Vlasov-Poisson systems will be utilized during the error analysis of each system. To proceed with the analysis, we will make assumptions about the existence of unique solutions and the function spaces these solutions come from. In all of the error estimate results, the above compatibility conditions will be assumed to hold. However, the DG methods that will be developed in this chapter will be seen to be well-defined under conditions that are less stringent than those given above.

4.4 DFUG method of approximation to the Vlasov equation

In this section, we consider the Vlasov system (4.34), where the data (α, f_0, f_I) is assumed, at first, to satisfy the Vlasov compatibility conditions. A new DG method, the *discontinuous flow upwind Galerkin* (DFUG) method, will be introduced to approximate this system. The main advantage of the DFUG method is that it is well-defined for given flows that are discontinuous. This is in contrast to the usual upwind Galerkin formulation for linear

transport equations, where the flow vector is required to be a continuous function over the domain Ω.

The motivation for designing a method for the Vlasov system that allows for broken flows comes from the fact that we ultimately want to design a DG method for the Vlasov-Poisson system. For this system, the potential, and hence the flow, are unknown and must be approximated, as well as the distribution f. The NIPG method will be used to approximate the potential. This implies that the gradient of the approximation of the potential will be broken on Ω^x, since ψ_h is itself broken. This then implies that the approximation to the flow α will be a broken function. Therefore, the DG method that we design for the Vlasov system will be such that it is defined for flows in the space $C^1(\,[0,T],\,[\,H^1_{div}(\mathcal{T}_h)\,]^6\,)$, where

$$[\,H^1_{div}(\mathcal{T}_h)\,]^6 \;=\; \{\,d(x,v) \in [\,H^1(\mathcal{T}_h)\,]^6 :\; \nabla \cdot d(x,v) \;=\; 0,\; \forall\; (x,v) \in \Omega\,\}\,,$$

even though the true potential is expected to be much smoother. If designed properly, the DFUG method should be easily incorporated into the overall DG method used to approximate the Vlasov-Poisson system.

Therefore, our first goal is derive a formulation for the Vlasov system defined by compatible data, which implies that the flow only satisfies the condition $C^1(\,[0,T],\,[\,H^1(\mathcal{T}_h)\,]^6\,)$. We want the derived formulation, which defines the DFUG method, to be such that if α satisfies the additional regularity that it is also in the space $C^1(\,[0,T],\,[\,C^1_{div}(\Omega)\,]^6\,)$, then the formulation reduces to the standard UG formulation [35],[29], which then implies that it is consistent, since the UG formulation results in a consistent DG method. We also want that the DFUG method is stable and is such that an $h-$optimal *a priori* error estimate can be established for this method.

4.4.1 Weak formulation

Our present objective is derive a weak formulation for the Vlasov system (4.34), (4.35), and (4.36) that is consistent and is defined even if α only has the regularity $C^1(\,[0,T],\,[\,H^1_{div}(\mathcal{T}_h)\,]^6\,)$. In order to ensure consistency, the following formulation will be derived by assuming that $\alpha \in C^1(\,[0,T],\,[\,C^1_{div}(\Omega)\,]^6\,)$, and then adjusting the formulation where appropriate so that it is well-defined for a less regular flow.

To begin the derivation, multiply equation (4.34) by an arbitrary test function $w \in H^1(\mathcal{T}_h)$ and integrate the resulting expression over an arbitrary element $K_j \in \mathcal{T}_h$. Then for each

$t \in (0,T]$ and for every element $K_j \in \mathcal{T}_h$ we get that

$$(f_t, w)_{K_j} - (\alpha f, \nabla w)_{K_j} + \langle f\,\alpha^- \cdot \nu_{K_j}, w^- \rangle_{\partial K_j/\Gamma} + \langle f\,\alpha \cdot \nu_{K_j}, w \rangle_{\partial K_j \cap \Gamma_O} = -\langle f_I\,\alpha \cdot \nu_{K_j}, w \rangle_{\partial K_j \cap \Gamma_I}, \tag{4.42}$$

$\forall\, w \in H^1(\mathcal{T}_h)$, where the notation α^- and w^- is introduced for simplicity and denotes the interior traces of α^- and w on K_j, i.e.,

$$w^-(x,v) = \lim_{s \to 0-} w\Big((x,v) + s\,\nu_{K_j}\Big) \qquad w^+(x,v) = \lim_{s \to 0+} w\Big((x,v) + s\,\nu_{K_j}\Big), \tag{4.43}$$

where ν_{K_j} is the outward unit normal to K_j, and likewise for α.

Upon summing the above equations over all elements it follows that, for each $t \in (0,T]$,

$$(f_t, w)_\Omega - \sum_{j=1}^{N_h} (\alpha f, \nabla w)_{K_j} + \sum_{j=1}^{N_h} \langle f\,\alpha^- \cdot \nu_{K_j}, w^- \rangle_{\partial K_j/\Gamma} + \sum_{f_k \in \Gamma_O} \langle f\,\alpha \cdot \nu_k, w \rangle_{f_k} = -\sum_{f_k \in \Gamma_I} \langle f_I\,\alpha \cdot \nu_k, w \rangle_{f_k}. \tag{4.44}$$

Using the fact that α is continuous on Ω, the above term containing integrations on $\partial K_j/\Gamma$ can be written as

$$\sum_{j=1}^{N_h} \langle f\,\alpha^- \cdot \nu_{K_j}, w^- \rangle_{\partial K_j/\Gamma} = \sum_{k=1}^{P_h} \langle f\,\alpha\,[\,w\,], \nu_k \rangle_{f_k} = \sum_{k=1}^{P_h} \langle f\,\overline{\alpha}\,[\,w\,], \nu_k \rangle_{f_k}, \tag{4.45}$$

where the average and jump operators, $\{\cdot\}$ and $[\cdot]$ are as defined in Chapter 2. Then (4.42) becomes

$$(f_t, w)_\Omega - \sum_{j=1}^{N_h} (\alpha f, \nabla w)_{K_j} + \sum_{k=1}^{P_h} \langle f\,\overline{\alpha}\,[\,w\,], \nu_k \rangle_{f_k} + \sum_{f_k \in \Gamma_O} \langle f\,\alpha \cdot \nu_k, w \rangle_{f_k} = -\sum_{f_k \in \Gamma_I} \langle f_I\,\alpha \cdot \nu_k, w \rangle_{f_k}. \tag{4.46}$$

We note that the above equation is now well-defined even if α is discontinuous across the interior faces of the mesh.

Discontinuous flow upwind function

Since our goal is to approximate f by a function F_h that is multi-valued on the interior faces of the mesh, equation (4.46) needs to be modified so that is well-defined even if f is discontinuous across the interior faces of the mesh. In particular, the terms involving f on the interior faces must be approximated. A standard technique is to replace f by its "upwind value" f^u on these faces, where the usual definition of f^u on an interior face $f_k = K_1 \cap K_2$, where $\nu_{K_1} = \nu_k$ and $\nu_{K_2} = -\nu_k$, is given by

$$f^u(\alpha)\,(x,v,t\,) = \begin{cases} f_{|K_1}(x,v,t\,) & , \quad \text{if } \alpha(x,t)\cdot\nu_k \geq 0\,, \\ f_{|K_2}(x,v,t\,) & , \quad \text{if } \alpha(x,t)\cdot\nu_k < 0\,. \end{cases} \tag{4.47}$$

Clearly, the definition is consistent in the sense that if f is continuous across an interior face then $f^u(\alpha) = f$ holds on this face. It is important to realize that the definition of f^u depends on the given flow field α. If α_1 and α_2 are two given flow fields having different values on some interior face f_k, then $f^u(\alpha_1) \neq f^u(\alpha_2)$ on f_k. In fact, f^u is not a linear function of α in the general case.

The definition given in (4.47) for the upwind function has the limitation that it only makes sense if α is continuous across all of the interior faces of the mesh. Since we want the DFUG method to be valid even when a flow is discontinuous at the interior faces, the definition in (4.47) must be modified. To fix this problem, we introduce the *discontinuous flow* upwind function f^u and define it to be

$$f^u(\alpha)\,(x,v,t\,) = \begin{cases} f_{|K_1}(x,v,t\,) & , \quad \text{if } \overline{\alpha}(x,t)\cdot\nu_k \geq 0\,, \\ f_{|K_2}(x,v,t\,) & , \quad \text{if } \overline{\alpha}(x,t)\cdot\nu_k < 0\,. \end{cases} \tag{4.48}$$

This definition is consistent with respect to the original upwind definition, in that it gives the same values on an interior face as the original function when α is continuous across this face. In the remainder of this work, the definition of f^u is taken to be that given in (4.48). For brevity, f^u is referred to as the upwind function, instead of the discontinuous flow upwind function.

We now replace f on the interior faces in (4.46) by its upwind value f^u. This leads to the

equation

$$(f_t, w)_\Omega - \sum_{j=1}^{N_h} (\alpha f, \nabla w)_{K_j} + \sum_{k=1}^{P_h} \langle f^u(\alpha)\,\overline{\alpha}\,[w], \nu_k \rangle_{f_k} + \sum_{f_k \in \Gamma_O} \langle f\alpha \cdot \nu_k, w \rangle_{f_k} = -\sum_{f_k \in \Gamma_I} \langle f_I\,\alpha \cdot \nu_k, w \rangle_{f_k}. \quad (4.49)$$

We remark that the above equation is now well-defined for functions f and w that are "broken" across the interior faces of the mesh $\mathcal{T}_h$.

Discontinuous flow stability term

The fact that the formulation being developed for the DFUG method allows for a discontinuous flow α poses a problem in terms of being able to show that the method is stable, for reasons that will soon be understood. To guarantee that the DFUG method is stable, equation (4.49) must be replaced by a slightly more complicated equation. In particular, the following stability term must be added to the formulation:

$$\frac{1}{2} \sum_{k=1}^{P_h} \langle \overline{f}\,[\alpha]\,[w], \nu_k \rangle_{f_k}. \quad (4.50)$$

We note that this term does not affect the consistency of the formulation, since it is equal to zero when the flow α is continuous.

Upon adding the above stability term to (4.49), we get the equation

$$(f_t, w)_\Omega - \sum_{j=1}^{N_h} (\alpha f, \nabla w)_{K_j} + \sum_{k=1}^{P_h} \langle f^u(\alpha)\,\overline{\alpha}\,[w] + \frac{1}{2}\overline{f}\,[\alpha]\,[w], \nu_k \rangle_{f_k} + \sum_{f_k \in \Gamma_O} \langle f\alpha \cdot \nu_k, w \rangle_{f_k} = -\sum_{f_k \in \Gamma_I} \langle f_I\,\alpha \cdot \nu_k, w \rangle_{f_k}. \quad (4.51)$$

Equation (4.51) is the weak formulation of the Vlasov equation (4.34) that is used to define the DFUG method. Some of the desirable properties of the formulation are that it is well-defined for functions f, α, and w that are discontinuous across the interior faces of the mesh, weakly enforces the inflow boundary condition (4.36), reduces to the standard UG formulation when α is continuous, and leads to a stable method. By imposing the boundary condition weakly, this condition does not need to be enforced through the definition of the

trial space of the DFUG method. This is a very valuable property to have if the method is to be numerically implemented, since one does not have to hassle with trying to enforce the boundary condition in some artificial manner.

4.4.2 Weak problem statement

We now state the definitions of a true and discrete solution to the DFUG formulation of the Vlasov system (4.34), (4.35), and (4.36). These definitions follow directly from the weak formulation (4.51).

Definition 13. [True DFUG Solution] *Given* $T > 0$ *and a compatible trio* (α, f_0, f_I), *a function* $f \in C^1([\,0,T\,],\, H^1(\mathcal{T}_h))$ *is said to be a true DFUG solution to the Vlasov system (4.34)-(4.36) if*

$$(i) \qquad f(\,t=0\,) \;=\; f_0 \qquad \textit{and} \tag{4.52}$$

$$\begin{aligned}(ii) \qquad & \forall\; t \in (\,0,T\,], \\ & (\,f_t, w\,)_\Omega \;-\; \sum_{j=1}^{N_h} (\,\alpha\, f, \nabla w\,)_{K_j} \;+\; \sum_{k=1}^{P_h} \langle\, f^u(\alpha)\, \overline{\alpha}\,[\,w\,] \;+\; \frac{1}{2}\,\overline{f}\,[\,\alpha\,]\,[\,w\,], \nu_k \,\rangle_{f_k} \\ & +\; \sum_{f_k \in \Gamma_O} \langle\, f\, \alpha \cdot \nu_{K_j}, w \,\rangle_{f_k} \;=\; -\sum_{f_k \in \Gamma_I} \langle\, f_I\, \alpha \cdot \nu_{K_j}, w \,\rangle_{f_k}\end{aligned} \tag{4.53}$$

is satisfied, $\forall\, w \in H^1(\mathcal{T}_h)$.

Similarly, the definition for the discrete case is as follows:

Definition 14. [Discrete DFUG Solution] *Given* $T > 0$ *and a compatible trio* (α, f_0, f_I), *a function* $F_h \in C^1([\,0,T\,],\, D_r(\mathcal{T}_h))$ *is said to be a discrete DFUG solution to the Vlasov system (4.34)-(4.36) if*

$$(i) \qquad (\,F_h(t=0), w\,)_{K_j} \;=\; (\,f_0, w\,)_{K_j}\,, \quad \forall\, K_j \in \mathcal{T}_h \qquad \textit{and} \tag{4.54}$$

$$\begin{aligned}(ii) \qquad & \forall\; t \in (\,0,T\,], \\ & (\,\frac{\partial}{\partial t} F_h, w\,)_\Omega \;-\; \sum_{j=1}^{N_h} (\,\alpha\, F_h, \nabla w\,)_{K_j} \;+\; \sum_{k=1}^{P_h} \langle\, (F_h)^u(\alpha)\, \overline{\alpha}\,[\,w\,] \;+\; \frac{1}{2}\,\bar{F}_h\,[\,\alpha\,]\,[\,w\,], \nu_k \,\rangle_{f_k} \\ & +\; \sum_{f_k \in \Gamma_O} \langle\, F_h\, \alpha \cdot \nu_{K_j}, w \,\rangle_{f_k} \;=\; -\sum_{f_k \in \Gamma_I} \langle\, f_I\, \alpha \cdot \nu_{K_j}, w \,\rangle_{f_k}\end{aligned} \tag{4.55}$$

is satisfied, $\forall\, w \in D_r(\mathcal{T}_h)$.

The existence and uniqueness of both a true and discrete DFUG solution will be discussed below. It will be seen that if there exists a true DFUG solution to the Vlasov system, then this solution is indeed unique. For the discrete case, it will be shown that there exists a unique solution satisfying (4.54) and (4.55).

4.4.3 Stability analysis

We now demonstrate that the DFUG method is stable. The stability result that is proved is such that it establishes that if a function f is a true DFUG solution to the Vlasov system, then it is a unique solution as well. By construction, a classical solution to the Vlasov system is a true DFUG solution, and hence, if such a solution exists, it is the unique true DFUG solution as well. As for the discrete case, the stability result can be used to show the existence and uniqueness of a discrete DFUG solution to the Vlasov system.

Before proceeding to the stability estimate, we first introduce a nonsymmetric, bilinear operator

$$b\,(\cdot,\cdot) : C^1(\,[\,0,T\,],\, H^1(\mathcal{T}_h)\,) \times C^1(\,[\,0,T\,],\, H^1(\mathcal{T}_h)\,) \to \mathbb{R}\,.$$

This operator is of fundamental importance when investigating both the stability and the *a priori* error estimate properties of the DFUG method. The definition given for b is as follows: $\forall\ (\,\xi, w\,) \in C^1(\,[\,0,T\,],\, H^1(\mathcal{T}_h)\,) \times C^1(\,[\,0,T\,],\, H^1(\mathcal{T}_h)\,)\,,$

$$\begin{aligned} b\,(\,\xi, w;\,\alpha\,) \;=\; & \int_0^T (\,\xi_t, w\,)_\Omega \;-\; \int_0^T \sum_{j=1}^{N_h} (\,\alpha\,\xi, \nabla w\,)_{K_j} \\ & +\; \int_0^T \sum_{k=1}^{P_h} \langle\, \xi^u(\alpha)\,\overline{\alpha}\,[\,w\,] \;+\; \frac{1}{2}\,\overline{\xi}\,[\,\alpha\,]\,[\,w\,], \nu_k \,\rangle_{f_k} \;+\; \sum_{f_k \in \Gamma_O} \int_0^T \langle\, \xi\,\alpha\cdot\nu_k, w\,\rangle_{f_k}\,. \end{aligned} \tag{4.56}$$

We note that the notation $b(\cdot,\cdot;\alpha)$ is used to reinforce that b depends explicitly on the flow α that defines the Vlasov system. Also, b can be thought of as a function of t, where $b(\cdot,\cdot;\alpha)(t)$ would be defined as above, except that the constant T would be replaced everywhere by $t \in [0,T]$. The reason that b plays such an important role in the upcoming analysis is now seen from the fact that if f is a true DFUG solution to the Vlasov system, then it follows

that

$$b\,(\,f, w;\,\alpha\,)\;=\;-\sum_{f_k\in\Gamma_I}\int_0^T\langle\,f_I\,\alpha\cdot\nu_k, w\,\rangle_{f_k} \tag{4.57}$$

holds, $\forall\, w\in C^1(\,[\,0,T\,],\,H^1(\mathcal{T}_h)$. Similarly, for the unique discrete DFUG solution F_h to the Vlasov system we have that

$$b\,(\,F_h, w;\,\alpha\,)\;=\;-\sum_{f_k\in\Gamma_I}\int_0^T\langle\,f_I\,\alpha\cdot\nu_k, w\,\rangle_{f_k} \tag{4.58}$$

holds, $\forall\, w\in C^1(\,[\,0,T\,],\,D^r(\mathcal{T}_h)\,)$.

The main property of b that will ultimately allow us to secure stability and error estimate results is now established. In proving the following lemma, it is important to note the role that the stability term (4.50) plays in obtaining the final result of the lemma.

Lemma 16. *Let* $\alpha\in C(\,[\,0,T\,],\,[\,H^1_{div}(\mathcal{T}_h)\,]^6\,)$. *Then the operator* b *satisfies the identity*

$$\begin{aligned}2\,b\,(\,\xi,\xi;\,\alpha\,)\;&=\;\|\,\xi(T)\,\|^2_{0,\Omega}\;+\;\int_0^T\sum_{k=1}^{P_h}\langle\,|\,\overline{\alpha}\cdot\nu_k\,|,[\,\xi\,]^2\,\rangle_{f_k}\\&+\;\int_0^T\sum_{f_k\in\Gamma}\langle\,|\,\alpha\cdot\nu_k\,|,\xi^2\,\rangle_{f_k}\;-\;\|\,\xi(0)\,\|^2_{0,\Omega}\,,\end{aligned} \tag{4.59}$$

$\forall\,\xi\in C^1(\,[\,0,T\,],\,H^1(\mathcal{T}_h)\,)$. *Moreover, identity (4.59) holds upon replacing* T *by any arbitrary* $t\in(0,T)$.

Proof. We will show that (4.59) holds at time T. In a similar manner, it can be shown that this identity holds for any arbitrary $t\in(0,T)$, as well.

Fix any arbitrary $\xi\in C^1(\,[\,0,T\,],\,H^1(\mathcal{T}_h)\,)$. Upon choosing $w=\xi$, one gets that

$$\begin{aligned}b\,(\,\xi,\xi;\,\alpha\,)\;&=\;\int_0^T(\,\xi_t,\xi\,)_\Omega\;-\;\int_0^T\sum_{j=1}^{N_h}(\,\alpha\,\xi,\nabla\xi\,)_{K_j}\\&+\;\int_0^T\sum_{k=1}^{P_h}\langle\,\xi^u(\alpha)\,\overline{\alpha}\,[\,\xi\,]\;+\;\frac{1}{2}\,[\,\alpha\,]\,\overline{\xi}\,[\,\xi\,],\nu_k\,\rangle_{f_k}\;+\;\int_0^T\sum_{f_k\in\Gamma_O}\langle\,\xi\,\alpha\cdot\nu_k,\xi\,\rangle_{f_k}\,.\end{aligned} \tag{4.60}$$

The proof now proceeds by writing each of the above righthandside terms in (4.61) in a

more suitable form.

We see that the first term satisfies

$$\int_0^T (\,\xi_t, \xi\,)_\Omega \;=\; \frac{1}{2}\,\|\,\xi(T)\,\|^2_{0,\Omega} \;-\; \frac{1}{2}\,\|\,\xi(0)\,\|^2_{0,\Omega}\,. \tag{4.61}$$

To simplify the second term, we integrate by parts and then take advantage of the divergence free nature of α to get

$$\begin{aligned}
-\int_0^T \sum_{j=1}^{N_h} (\,\alpha\xi, \nabla\xi\,)_{K_j} \;&=\; -\int_0^T \sum_{j=1}^{N_h} (\,\alpha, \frac{1}{2}\,\nabla(\xi^2)\,)_{K_j} \;=\; -\frac{1}{2}\int_0^T \sum_{j=1}^{N_h} \langle\,\alpha^-\,(\xi^-)^2, \nu_{K_j}\,\rangle_{\partial K_j} \\
&=\; -\frac{1}{2}\int_0^T \sum_{j=1}^{N_h} \Big(\langle\,\alpha^-\,(\xi^-)^2, \nu_{K_j}\,\rangle_{\partial K_j/\Gamma} \;+\; \langle\,\alpha\,\xi^2, \nu\,\rangle_{\partial K_j\cap\Gamma_O} \;+\; \langle\,\alpha\,\xi^2, \nu\,\rangle_{\partial K_j\cap\Gamma_I} \Big) \\
&=\; -\frac{1}{2}\int_0^T \left(\sum_{k=1}^{P_h} \langle\,[\,\alpha\,\xi^2\,], \nu_k\,\rangle_{f_k} \;+\; \sum_{f_k\in\Gamma_O} \langle\,\alpha\,\xi^2, \nu_k\,\rangle_{f_k} \;+\; \sum_{f_k\in\Gamma_I} \langle\,\alpha\,\xi^2, \nu_k\,\rangle_{f_k} \right).
\end{aligned} \tag{4.62}$$

To simplify the above term involving integrations over the interior faces, we use the fact that

$$\begin{aligned}
[\,\alpha\,\xi^2\,] \;&=\; \alpha_1\xi_1^2 \;-\; \alpha_2\xi_2^2 \;=\; (\,\alpha_1 \;+\; \alpha_2\,)(\,\xi_1^2 \;-\; \xi_2^2\,) \;+\; \alpha_1\xi_2^2 \;-\; \alpha_2\xi_1^2 \\
&=\; 2\,\overline{\alpha}\,[\,\xi^2\,] \;+\; \alpha_1\xi_2^2 \;-\; \alpha_2\xi_2^2 \;+\; \alpha_2\xi_2^2 \;-\; \alpha_1\xi_1^2 \;+\; \alpha_1\xi_1^2 \;-\; \alpha_2\xi_1^2 \\
&=\; 2\,\overline{\alpha}\,[\,\xi^2\,] \;+\; [\,\alpha\,]\xi_2^2 \;-\; [\,\alpha\,\xi^2\,] \;+\; [\,\alpha\,]\xi_1^2 \\
&=\; 2\,\overline{\alpha}\,[\,\xi^2\,] \;+\; [\,\alpha\,](\,\xi_1^2 \;+\; \xi_2^2\,) \;-\; [\,\alpha\,\xi^2\,] \\
&=\; 2\,\overline{\alpha}\,[\,\xi^2\,] \;+\; 2\,[\,\alpha\,]\,\overline{\xi}\,[\,\xi\,] \;-\; [\,\alpha\,\xi^2\,],
\end{aligned} \tag{4.63}$$

which implies

$$[\,\alpha\,\xi^2\,] \;=\; \overline{\alpha}\,[\,\xi^2\,] \;+\; [\,\alpha\,]\,\overline{\xi}\,[\,\xi\,]\,. \tag{4.64}$$

So, identity (4.64) implies that (4.62) can be rewritten as

$$
\begin{aligned}
-\int_0^T \sum_{j=1}^{N_h} (\,\alpha\xi, \nabla\xi\,)_{K_j} \;&=\; -\frac{1}{2}\int_0^T \sum_{k=1}^{P_h} \langle\, \overline{\alpha}\,[\,\xi^2\,] \;+\; [\,\alpha\,]\,\overline{\xi}\,[\,\xi\,], \nu_k \,\rangle_{f_k} \\
&-\; \frac{1}{2}\int_0^T \left(\sum_{f_k \in \Gamma_O} \langle\, \alpha\,\xi^2, \nu_k \,\rangle_{f_k} \;+\; \sum_{f_k \in \Gamma_I} \langle\, \alpha\,\xi^2, \nu_k \,\rangle_{f_k} \right). \qquad (4.65)
\end{aligned}
$$

Upon inserting (4.65) and (4.61) into (4.60), we see that $b\,(\xi, \xi; \alpha)$ can be written as

$$
\begin{aligned}
b\,(\,\xi, \xi;\, \alpha\,) \;&=\; \frac{1}{2}\,\|\,\xi(T)\,\|^2_{0,\Omega} \;+\; \int_0^T \sum_{k=1}^{P_h} \langle\, \xi^u(\alpha)\,\overline{\alpha}\,[\,\xi\,] \;-\; \frac{1}{2}\,\overline{\alpha}\,[\,\xi^2\,], \nu_k \,\rangle_{f_k} \\
&+\; \frac{1}{2}\int_0^T \sum_{f_k \in \Gamma_O} \langle\, \xi\,\alpha\cdot\nu_k, \xi \,\rangle_{f_k} \;-\; \frac{1}{2}\int_0^T \sum_{f_k \in \Gamma_I} \langle\, \xi\,\alpha\cdot\nu_k, \xi \,\rangle_{f_k} \;-\; \frac{1}{2}\,\|\,\xi(0)\,\|^2_{0,\Omega}\,. \qquad (4.66)
\end{aligned}
$$

The reason for adding the stability term is now clear from the above equality. By adding this term, we were able to cancel the term $-\frac{1}{2}\,[\,\alpha\,]\,\overline{\xi}\,[\,\xi\,]$ in the righthandside of (4.64). After performing this cancelation, the remainder of the stability proof proceeds in a manner similar to the case when α is a given smooth function. The point here is that if one allows the flow α to be discontinous, then one must add an appropriate term to the weak formulation to maintain stability.

We now continue by manipulating each of the interior face terms in (4.66) as follows:

$$
\begin{aligned}
\sum_{k=1}^{P_h} \langle\, \xi^u(\alpha)\,\overline{\alpha}\,[\,\xi\,] \;-\; \frac{1}{2}\,\overline{\alpha}\,[\,\xi^2\,], \nu_k \,\rangle_{f_k} \;&=\; \sum_{k=1}^{P_h} \langle\, \xi^u(\alpha)\,\overline{\alpha}\,[\,\xi\,] \;-\; \overline{\alpha}\,\overline{\xi}\,[\,\xi\,], \nu_k \,\rangle_{f_k} \\
&=\; \sum_{k=1}^{P_h} \langle\, (\,\xi^u(\alpha) \;-\; \overline{\xi}\,)\,\overline{\alpha}\cdot\nu_k, [\,\xi\,] \,\rangle_{f_k}\,, \qquad (4.67)
\end{aligned}
$$

where the last line follows from the fact that $\frac{1}{2}\,[\,\xi^2\,] \;=\; \bar{\xi}\;[\,\xi\,]$.

To further simplify (4.67), we appeal to the definition of f^u. If $\overline{\alpha}\cdot\nu_k \geq 0$ holds, then we have that $\xi^u(\alpha) = \xi_1$. This, in turn, implies that

$$
(\,\xi^u(\alpha) \;-\; \overline{\xi}\,)\,\overline{\alpha}\cdot\nu_k \;=\; (\,\frac{1}{2}\,\xi_1 \;-\; \frac{1}{2}\,\xi_2\,)\,\overline{\alpha}\cdot\nu_k \;=\; \frac{1}{2}\,[\,\xi\,]\,\overline{\alpha}\cdot\nu_k \;=\; \frac{1}{2}\,[\,\xi\,]\,|\,\overline{\alpha}\cdot\nu_k\,|\,. \qquad (4.68)
$$

If we instead have $\overline{\alpha}\cdot\nu_{_k} < 0$, then it follows that $\xi^u(\alpha) = \xi_{_2}$. This implies

$$(\,\xi^u(\alpha)\;-\;\overline{\xi}\,)\,\overline{\alpha}\cdot\nu_{_k}\;=\;(\,\frac{1}{2}\,\xi_{_2}\;-\;\frac{1}{2}\,\xi_{_1}\,)\,\overline{\alpha}\cdot\nu_{_k}\;=\;-\frac{1}{2}\,[\,\xi\,]\,\overline{\alpha}\cdot\nu_{_k}\;=\;\frac{1}{2}\,[\,\xi\,]\,|\,\overline{\alpha}\cdot\nu_{_k}\,|\,. \quad (4.69)$$

Thus, (4.67) can be rewritten as

$$\begin{aligned}\sum_{k=1}^{P_h}\langle\,\xi^u(\alpha)\,\overline{\alpha}\,[\,\xi\,]\;-\;\frac{1}{2}\,\overline{\alpha}\,[\,\xi^2\,],\nu_{_k}\,\rangle_{_{f_k}}\;&=\;\sum_{k=1}^{P_h}\langle\,(\,\xi^u(\alpha)\;-\;\overline{\xi}\,)\,\overline{\alpha}\cdot\nu_{_k},[\,\xi\,]\,\rangle_{_{f_k}}\\ &=\;\frac{1}{2}\sum_{k=1}^{P_h}\langle\,|\,\overline{\alpha}\cdot\nu_{_k}\,|,[\,\xi\,]^2\,\rangle_{_{f_k}}\;\geq 0\,. \quad (4.70)\end{aligned}$$

This shows that the summation of all of the interior face terms in (4.66) is a nonnegative quantity.

In a similar manner, we can show the boundary terms in (4.66) satisfy

$$\sum_{f_k\in\Gamma_{_O}}\langle\,\xi\,\alpha\cdot\nu_{_k},\xi\,\rangle_{_{f_k}}\;=\;\sum_{f_k\in\Gamma_{_I}}\langle\,|\,\alpha\cdot\nu_{_k}\,|,\xi^2\,\rangle_{_{f_k}}\qquad\text{and} \quad (4.71)$$

$$-\sum_{f_k\in\Gamma_{_I}}\langle\,\xi\,\alpha\cdot\nu_{_k},\xi\,\rangle_{_{f_k}}\;=\;\sum_{f_k\in\Gamma_{_I}}\langle\,|\,\alpha\cdot\nu_{_k}\,|,\xi^2\,\rangle_{_{f_k}}\,, \quad (4.72)$$

which shows that the summation of all of the boundary terms in (4.66) is a nonnegative quantity.

Upon inserting equations (4.70), (4.71), and (4.72) into (4.66), we obtain identity (4.59). □

We now present the stability result for DFUG method. The inequality established demonstrates that if f is a true DFUG solution to the Vlasov system, then it is controlled by three functions: the initial condition $f_{_0}$, the inflow boundary condition $f_{_I}$, and the flow α. It will be seen that in proving the following lemma, explicit use is made of the identity for b given in Lemma 16.

Lemma 17. [Stability of the DFUG method] *If f is a true DFUG solution to the Vlasov*

sytem, then

$$\| f(T) \|_{0,\Omega}^2 + \frac{1}{2} \int_0^T \left(\sum_{k=1}^{P_h} \| \, | \, \overline{\alpha} \cdot \nu_k \, |^{1/2} \, [\, f \,] \, \|_{0,f_k}^2 + \sum_{f_k \in \Gamma} \| \, | \, \alpha \cdot \nu_k \, |^{1/2} \, f \, \|_{0,f_k}^2 \right)$$
$$\leq \| f_0 \|_{0,\Omega}^2 + 2 \int_0^T \sum_{f_k \in \Gamma_I} \| \, | \, \alpha \cdot \nu_k \, |^{1/2} \, f_I^2 \, \|_{0,f_k} . \tag{4.73}$$

Proof. Since f is a true DFUG solution to the Vlasov system, it satisfies $f(t=0) = f_0$ and

$$b(f, f; \alpha) = \int_0^T \sum_{f_k \in \Gamma_I} \langle \, f_I \, | \, \alpha \cdot \nu_k \, |, f \, \rangle_{f_k} . \tag{4.74}$$

Applying Lemma 16, we then see that f satisfies

$$\| f(T) \|_{0,\Omega}^2 + \int_0^T \sum_{k=1}^{P_h} \langle \, | \, \overline{\alpha} \cdot \nu_k \, |, [\, f \,]^2 \, \rangle_{f_k} + \int_0^T \sum_{f_k \in \Gamma_O} \langle \, | \, \alpha \cdot \nu_k \, |, f^2 \, \rangle_{f_k}$$
$$+ \int_0^T \sum_{f_k \in \Gamma_I} \langle \, | \, \alpha \cdot \nu_k \, |, f^2 \, \rangle_{f_k} = \| f_0 \|_{0,\Omega}^2 + 2 \int_0^T \sum_{f_k \in \Gamma_I} \langle \, f_I \, | \, \alpha \cdot \nu_k \, |, f \, \rangle_{f_k} . \tag{4.75}$$

To bound the term involving the inflow boundary condition in the above equation, we proceed by noting that

$$2 \int_0^T \sum_{f_k \in \Gamma_I} \langle \, f_I \, | \, \alpha \cdot \nu_k \, |, f \, \rangle_{f_k}$$
$$\leq 2 \int_0^T \sum_{f_k \in \Gamma_I} \langle \, f_I \, | \, \alpha \cdot \nu_k \, |, f_I \, \rangle_{f_k} + \frac{1}{2} \int_0^T \sum_{f_k \in \Gamma_I} \langle \, f \, | \, \alpha \cdot \nu_k \, |, f \, \rangle_{f_k} . \tag{4.76}$$

Thus, the above inequality and (4.75) together imply that

$$\| f(T) \|_{0,\Omega}^2 + \int_0^T \Big(\sum_{k=1}^{P_h} \| \, | \, \overline{\alpha} \cdot \nu_k \, |^{1/2} \, [\, f \,] \, \|_{0,f_k}^2 + \sum_{f_k \in \Gamma_O} \| \, | \, \alpha \cdot \nu_k \, |^{1/2} \, f \, \|_{0,f_k}^2$$
$$+ \frac{1}{2} \sum_{f_k \in \Gamma_I} \| \, | \, \alpha \cdot \nu_k \, |^{1/2} \, f \, \|_{0,f_k}^2 \Big) \leq \| f_0 \|_{0,\Omega}^2 + 2 \int_0^T \sum_{f_k \in \Gamma_I} \| \, | \, \alpha \cdot \nu_k \, |^{1/2} \, f_I \, \|_{0,f_k}^2 . \tag{4.77}$$

This clearly implies the stability inequality (4.73), since the lefthandside in (4.77) bounds

the lefthandside in (4.73). □

The uniqueness of a true DFUG solution to the Vlasov system is now easy to deduce from Lemma 16 and the linearity of $b\,(\cdot, w; \alpha)$.

Corollary 1. [Uniquess of the DFUG method] *If f is a true DFUG solution to the Vlasov system, then it is unique.*

Proof. For a given flow α, suppose that f_1 and f_2 are both DFUG solutions. Since the DFUG method is linear in f, it follows that

$$b\,(\,f_1 - f_2, w;\,\alpha\,)\;=\;0\,, \tag{4.78}$$

$\forall\; t \in (\,0, T\,],\;\forall\; w \in C^1(\,[\,0,T\,],\,H^1(\mathcal{T}_h)\,)$. Setting $w = f_1 - f_2$ in the above equation and then using identity (4.59), and the fact that $f_1(t=0) = f_2(t=0)$ we find that

$$\begin{aligned}\|\,(f_1 - f_2)\,(s)\,\|^2_{0,\Omega}\;+\;\int_0^s \sum_{k=1}^{P_h} \langle\,|\,\overline{\alpha}\cdot\nu_k\,|, [\,f_1 - f_2\,]^2\,\rangle_{f_k}\;+\;\int_0^s \sum_{f_k\in\Gamma_O} \langle\,|\,\alpha\cdot\nu_k|, (f_1-f_2)^2\,\rangle_{f_k}\\ +\;\int_0^s \sum_{f_k\in\Gamma_O} \langle\,|\,\alpha\cdot\nu_k\,|, (f_1-f_2)^2\,\rangle_{f_k}\;=\;0\,,\end{aligned} \tag{4.79}$$

holds $\forall\; s \in (\,0,T\,]$. Thus, $f_1 \equiv f_2$. □

The existence of a true DFUG solution to the Vlasov system is guaranteed if the Vlasov system admits a classical solution. This results from the fact that the DFUG method is a consistent method by design, i.e., if f is a classical solution of the Vlasov system, then it is also a DFUG solution and if f is a DFUG solution and is smooth enough, then it is a classical solution as well.

The solvability of the discrete Vlasov problem is stated in the following result.

Lemma 18. [Existence and uniquess of the discrete DFUG method] *There exists a unique discrete DFUG solution F_h to the Vlasov system. Moreover, F_h satisfies the stability in-*

equality

$$\| F_h(T) \|_{0,\Omega}^2 + \frac{1}{2} \int_0^T \left(\sum_{k=1}^{P_h} \| \, | \, \overline{\alpha} \cdot \nu_k \, |^{1/2} \, [\, F_h \,] \, \|_{0,f_k}^2 + \| \, | \, \alpha \cdot \nu_k \, |^{1/2} \, F_h \, \|_{0,\Gamma}^2 \right)$$
$$\leq \| f_0 \|_{0,\Omega}^2 + 2 \int_0^T \| \, | \, \alpha \cdot \nu_k \, |^{1/2} f_I^2 \, \|_{0,\Gamma_I} \, . \tag{4.80}$$

Proof. Inequality (4.80) is proved in the same way as (4.73) was proved in Lemma 17. Since the discrete DFUG Vlasov problem is linear and finite dimensional, existence and uniqueness are equivalent. Since (4.80) implies any discrete DFUG solution is unique, it also implies the existence of a discrete solution. □

4.5 DFUG approximation to the perturbed flow Vlasov equation

We now consider using the DFUG method to approximate a perturbed flow Vlasov system. A perturbed flow Vlasov system is a Vlasov system defined by a compatible trio $(\aleph_h, f_0, f_I)$, where $\aleph_h$ is interpreted as some perturbation of the flow α. Throughout this section, the trio (α, f_0, f_I) will be assumed to be compatible according to Definition 12, which implies $\alpha \in C^1(\, [0,T], \, [\, C^1_{div}(\Omega) \,]^6 \,)$.

In the remainder of this section, we will assume that $\{ \, \aleph_h \, \}_{h>0}$ is a sequence of perturbed flows defined by an underlying sequence of perturbed potentials $\{ \, \psi_h \, \}_{h>0}$ satisfying $\{ \, \psi_h \, \}_{h>0} \subset C^1(\, [0,T], \, H^2(\mathcal{T}_h) \,), \, \forall \, h > 0$, which then implies that $\aleph_h \in C^1(\, [0,T], \, [\, H^1_{div}(\mathcal{T}_h) \,]^6 \,), \, \forall \, h > 0$.

The objective of this section is to derive an *a priori* error estimate that quantifies the error between the true DFUG solution to the Vlasov system defined by the (α, f_0, f_I) and the discrete DFUG solution to the perturbed Vlasov system defined by the trio $(\aleph_h, f_0, f_I)$. This estimate will be seen to contain contributions arising from the DFUG discretization of the Vlasov system and will also contain contributions coming from the difference between α and $\aleph_h$. In particular, we will establish an error estimate in Theorem 14 that depends explicitly on the quantities

$$\begin{array}{lll} (i) & \| \nabla_x (\psi - \psi_h) \|_{0,\Omega^x} \, , & \\ (ii) & \| \, [\, \nabla_x \psi_h \,] \, \|_{0,\mathcal{G}_h^x} \, , & \text{and} \\ (iii) & \| \nabla_x (\psi - \psi_h) \|_{0,\partial\Omega^x} \, . & \end{array}$$

Thus, the results of Theorem 14 will show exactly how the perturbation errors in the flow effect the approximation error between the true solution to the Vlasov system and the discrete DFUG solution to the perturbed Vlasov system.

To avoid any ambiguity, we now state what it precisely means for given data to the perturbed Vlasov problem to be compatible, which we will refer to as perturbed compatibility, and to be a discrete DFUG solution to a perturbed Vlasov system.

Definition 15. [Data Compatibility for Perturbed Vlasov System] *Given $T > 0$, the data trio $(\aleph_h, f_0, f_I)$ is said to be perturbed compatible up to time T if*

$$
\begin{aligned}
f_0 &\in C^1(\,\Omega^x,\, C_c^1(\Omega^v)\,)\,, && (4.81)\\
f_I &\in C^1(\,\partial\Omega \times [0,T]\,)\,, && (4.82)\\
f_I(\cdot, v, \cdot) &= 0\,, \quad \forall\ v \in \partial\Omega^v\,, && (4.83)\\
\partial^{|\,\beta\,|} f_I(t=0) &= \partial^{|\,\beta\,|} f_0\,, \quad \text{on } \Gamma_I(t=0),\ \forall\ |\,\beta\,| \ \le\ 1\,, \text{ and} && (4.84)\\
\psi_h &\in C^1(\,[0,T],\, H^2(\Omega^x)\,)\,, && (4.85)
\end{aligned}
$$

where

$$
\aleph_h(x,v,t) \ =\ \begin{pmatrix} v \\ \nabla_x \psi_h(x,t) \end{pmatrix}\,. \tag{4.86}
$$

So, if $(\aleph_h, f_0, f_I)$ is perturbed compatible, then $\aleph_h \in C^1(\,[0,T],\, [\,H^1_{div}(\Omega)\,]^6\,)$. The definition of a discrete DFUG solution to the perturbed Vlasov system can now be defined.

Definition 16. [Discrete DFUG Solution to perturbed Vlasov System] *Given $T > 0$ and a perturbed compatible trio $(\aleph_h, f_0, f_I)$, a function $F_h \in C^1(\,[\,0,T\,],\, D_r(\mathcal{T}_h)\,)$ is said to be a discrete DFUG solution to the perturbed Vlasov system if*

$$
(i) \qquad (\,F_h(t=0), w\,)_{K_j} \ =\ (\,f_0, w\,)_{K_j}\,, \quad \forall\, K_j \in \mathcal{T}_h \qquad \textit{and} \tag{4.87}
$$

$$
\begin{aligned}
(ii) \qquad & \forall\ t \in (\,0,T\,]\,,\\
& (\,\frac{\partial}{\partial t} F_h, w\,)_\Omega \ -\ \sum_{j=1}^{N_h} (\,\aleph_h\, F_h, \nabla w\,)_{K_j} \ +\ \sum_{k=1}^{P_h} \langle\, (F_h)^u(\aleph_h)\, \overline{\aleph_h}\,[\,w\,] \ +\ \frac{1}{2}\, \bar{F}_h\,[\,\aleph_h\,]\,[\,w\,], \nu_k \,\rangle_{f_k}\\
& +\ \sum_{f_k \in \Gamma_O} \langle\, F_h\, \aleph_h \cdot \nu_{K_j}, w \,\rangle_{f_k} \ =\ -\ \sum_{f_k \in \Gamma_I} \langle\, f_I\, \aleph_h \cdot \nu_{K_j}, w \,\rangle_{f_k}
\end{aligned} \tag{4.88}
$$

is satisfied, $\forall\, w \in D_r(\mathcal{T}_h)$.

The solvability of the approximate discrete Vlasov problem is stated in the following result.

Lemma 19. [Existence & Uniquess of Discrete DFUG Solution to Perturbed Vlasov System] *There exists a unique discrete DFUG solution F_h to the perturbed Vlasov system. Moreover, F_h satisfies the stability inequality*

$$\| F_h(T) \|_{0,\Omega}^2 + \frac{1}{2} \int_0^T \left(\sum_{k=1}^{P_h} \| \, | \overline{\aleph_h} \cdot \nu_k \, |^{1/2} \, [F_h] \, \|_{0,f_k}^2 + \sum_{f_k \in \Gamma_O} \| \, | \aleph_h \cdot \nu_k \, |^{1/2} \, F_h \, \|_{0,\Gamma}^2 \right)$$
$$\leq \| f_0 \|_{0,\Omega}^2 + 2 \int_0^T \sum_{f_k \in \Gamma_I} \| \, | \aleph_h \cdot \nu_k \, |^{1/2} f_I^2 \, \|_{0,\Gamma_I} . \tag{4.89}$$

Proof. The proof is exactly the same as that given for the discrete DFUG solution in Lemma 19. □

Throughout the rest of this section, we will use F_h to denote the discrete DFUG solution to the perturbed Vlasov system defined by $\aleph_h$, and not the discrete DFUG solution to the Vlasov system defined by α.

4.5.1 Pseudo-Galerkin orthogonality

To appreciate the complication that arises when trying to estimate the error between the true DFUG solution to the Vlasov system and the discrete DFUG solution to the perturbed Vlasov system, we make a few observations. First, since f is a solution to the Vlasov system defined by α, it satisfies the equation

$$b(f, w; \alpha) = -\langle f_I \, \alpha \cdot \nu_k, w \rangle_{\Gamma_I(\alpha)(t)} , \qquad \forall \, w \in H^1(\mathcal{T}_h) . \tag{4.90}$$

In contrast, since the F_h is a discrete solution to the perturbed Vlasov system defined by $\aleph_h$, instead of α, it satisfies the equation

$$b(F_h, w; \aleph_h) = -\langle f_I \, \aleph_h \cdot \nu_k, w \rangle_{\Gamma_I(\aleph_h)(t)} , \qquad \forall \, w \in D^r(\mathcal{T}_h) . \tag{4.91}$$

Although, the above two equations for the true and discrete solutions appear similar, there is a subtle, but significant difference. To see this, consider the discrete DFUG solution f_h

to the Vlasov system. By definition, f_h satisfies

$$b(\, f_h,\, w;\, \alpha\,) \;=\; -\langle\, f_{_I}\, \alpha \cdot \nu_{_k}, w\,\rangle_{\Gamma_{_I}(\alpha)(t)}\,, \qquad \forall\; w \in D^r(\mathcal{T}_h)\,. \tag{4.92}$$

Upon combining (4.90) and (4.92) together, we would get the following standard Galerkin orthogonality relationship:

$$b(\, f_h,\, w;\, \alpha\,) \;=\; b(\, f,\, w;\, \alpha\,)\,, \qquad \forall\; w \in D^r(\mathcal{T}_h)\,. \tag{4.93}$$

To find a similar relationship that holds between F_h and f, we make use of the fact that $f_{_I} = 0$ on $\Omega^x \times \partial\Omega^v$ and that $\aleph_h \cdot \nu = \alpha \cdot \nu$ on $\partial\Omega^x \times \Omega^v$, which follows from the fact that $\Omega^x = [0, L_1] \times [0, L_2] \times [0, L_3]$. These facts taken together imply that the righthandside of (4.91) is equal to the righthandside of (4.90). Hence, the following identity holds:

$$b\,(F_h, w; \aleph_h) \;=\; b(\, f,\, w;\, \alpha\,)\,, \qquad \forall\; w \in D^r(\mathcal{T}_h)\,. \tag{4.94}$$

This relationship is referred to as the *pseudo-Galerkin* orthogonality relationship, since it is the closest relationship to Galerkin orthogonality between the functions f and F_h.

It is worthwhile to mention that if the above simplifications that are assumed to hold on $\partial\Omega = (\partial\Omega^x \times \Omega^v) \times (\partial\Omega^x \times \Omega^v)$ do not hold, then (4.94) would not be true. Instead, we would get the relationship

$$b\,(F_h, w; \aleph_h) \;=\; b(\, f,\, w;\, \alpha\,) \;+\; \langle\, f_{_I}\, \alpha \cdot \nu_{_k}, w\,\rangle_{\Gamma_{_I}(\alpha)(t)} \;-\; \langle\, f_{_I}\, \aleph_h \cdot \nu_{_k}, w\,\rangle_{\Gamma_{_I}(\aleph_h)(t)}\,,\; \forall\; w \in D^r(\mathcal{T}_h)\,. \tag{4.95}$$

In this case, the following error analysis would still be valid, but additional work would be required to bound the difference between the last two boundary terms in (4.95).

4.5.2 *A priori* error analysis

In this section, we derive *a priori* error estimates between the true solution to the Vlasov system and the discrete DFUG solution to the perturbed Vlasov system.

To begin the analysis, we first see what quantitities are natural to estimate in the DFUG

formulation. From (4.59), we have that

$$
\begin{aligned}
b\,(\,f-F_h, f-F_h;\,\aleph_h\,)\;+\;\|\,f(0)-F_h(0)\,\|^2_{0,\Omega}\;&=\;\|\,f(T)-F_h(T)\,\|^2_{0,\Omega}\\
&+\;\int_0^T\sum_{k=1}^{P_h}\|\,|\,\overline{\aleph_h}\cdot\nu_{_k}\,|^{1/2}\,[\,f-F_h\,]\,\|^2_{0,f_k}\;+\;\int_0^T\|\,|\,\aleph_h\cdot\nu_{_k}\,|^{1/2}\,[\,f-F_h\,]\,\|^2_{0,\Gamma_{_O}}\\
&+\;\int_0^T\|\,|\,\aleph_h\cdot\nu_{_k}\,|^{1/2}\,[\,f-F_h\,]\,\|^2_{0,\Gamma_{_I}}\,. \qquad (4.96)
\end{aligned}
$$

Thus, we see that the formulation is suitable for trying to establish error bounds for the three quantities in the above righthandside. In fact, the estimate we derive for the above quantities will not just be valid for time T, but the estimate will hold for any $t\in[\,0,T\,]$. We note that these three quantities do not depend explicitly on α but only on $\aleph_h$. The only dependence on α is through the function f.

The error estimate for the DFUG method is now established. It will be seen that the following theorem gives an optimal estimate in h when the perturbed flow is set equal α. In the case when the perturbed flow is different from α, it cannot be known whether the estimate is optimal or not, since this depends on how well the sequence $\{\aleph_h\}$ approximates α.

Theorem 14. *Let $(\alpha, f_{_0}, f_{_I})$ be a compatible trio. Assume that there exists a unique classical solution f to the Vlasov system defined by $(\alpha, f_{_0}, f_{_I})$, where $f\in C^1([0,T],H^s(\mathcal{T}_h))$, $s>3$, is satisfied. Let $(\aleph_h, f_{_0}, f_{_I})$ be a perturbed compatible trio and let $F_h\in D^r(\mathcal{T}_h)$ be the discrete DFUG solution to the perturbed Vlasov system defined by $(\aleph_h, f_{_0}, f_{_I})$. Then $\exists$ a mesh independent constant C that scales linearly with the final time T such that the following a priori error estimate holds:*

$$
\begin{aligned}
&\|\,f(T)-F_h(T)\,\|^2_{0,\Omega}\;+\;\int_0^T|\!|\!|\,|\,\overline{\aleph_h}\cdot\nu_{_k}\,|^{1/2}\,[\,f-F_h\,]\,|\!|\!|^2_{0,\hat{\mathscr{F}}_h}\;+\;\int_0^T|\!|\!|\,|\,\aleph_h\cdot\nu_{_k}\,|^{1/2}\,(\,f-F_h\,)\,|\!|\!|^2_{0,\Gamma_{_O}}\\
&+\;\int_0^T|\!|\!|\,|\,\aleph_h\cdot\nu_{_k}\,|^{1/2}\,(\,f-F_h\,)\,|\!|\!|^2_{0,\Gamma_{_I}}\\
&\leq\;C\left(h^{2\mu-1}\;+\;h^{2\mu+2s-8}\int_0^T|\!|\!|\,\nabla_x(\psi-\psi_h)\,|\!|\!|_{0,\Omega^x}\right)\\
&+\;\int_0^T\left(h^{2\mu+2s-8}\,|\!|\!|\,\nabla_x(\psi-\psi_h)\,|\!|\!|^2_{0,\Omega^x}\;+\;h^{2\mu+2s-7}|\!|\!|\,[\,\nabla_x\psi_h\,]\,|\!|\!|^2_{0,\hat{\mathcal{G}}_h^x}\right.\\
&\left.+\;h^{-1}\,|\!|\!|\,[\,\nabla_x\psi_h\,]\,|\!|\!|^2_{0,\hat{\mathcal{G}}_h^x}\;+\;h^{-1}\,|\!|\!|\,\nabla_x(\psi-\psi_h)\,|\!|\!|^2_{0,\partial\Omega^x}\right), \qquad (4.97)
\end{aligned}
$$

where

$$\mu \;=\; \min\{r+1, s\}\,.$$

Moreover, the above estimate holds upon replacing T everywhere by any fixed $t \in [\,0, T\,]$. Also, upon setting $\aleph_h = \alpha$, the above estimate reduces to

$$\| f(T) - F_h(T) \|_{0,\Omega}^2 \;+\; \int_0^T \sum_{k=1}^{P_h} \| \, | \overline{\aleph_h} \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,f_k}^2 \;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,\Gamma_O}^2$$
$$+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,\Gamma_I}^2 \;\leq\; C\, h^{2\mu - 1}\,, \tag{4.98}$$

which is optimal in h.

Proof. Since f is a classical solution to the Vlasov-Poisson system defined by (α, f_0, f_I), it is also a unique, true DFUG solution to this same system, which follows by the stability of the DFUG method.

To begin the proof, we slpit the estimation problem into two seperate problems, one that depends only on $f - \Pi_h^r f$ and one that depends only on $F_h - \Pi_h^r f$.

To split the problem, we apply the triangle inequality and Young's inequality:

$$\begin{aligned} b\,(\, f - F_h, f - F_h; \aleph_h\,) \;+\; \| \, f_0 - F_h(0) \, \|_{0,\Omega}^2 & \\ \leq\; 2\, \Big(b(\, f - \Pi_h^r f, f - \Pi_h^r f; \aleph_h\,) \;+\; \| \, f_0 - \Pi_h^r f(0) \, \|_{0,\Omega}^2 \Big) & \\ +\; 2\, \Big(b(\, F_h - \Pi_h^r f, F_h - \Pi_h^r f; \aleph_h\,) \;+\; \| \, F_h(0) - \Pi_h^r f(0) \, \|_{0,\Omega}^2 \Big) & \\ =\; 2\, \Big(b(\, f - \Pi_h^r f, f - \Pi_h^r f; \aleph_h\,) \;+\; \| \, f_0 - \Pi_h^r f(0) \, \|_{0,\Omega}^2 \Big) & \\ +\; 2\, b\,(\, F_h - \Pi_h^r f, F_h - \Pi_h^r f; \aleph_h\,)\,, & \end{aligned} \tag{4.99}$$

where the last line follows from the fact that $F_h(t=0) \;=\; \Pi_h^r f(t=0)\,.$

We now continue by estimating the first two terms in the righthandside of (4.99). By the interpolation properties from Theorem 7 that hold for $\Pi_h^r f$ and by similar estimates to (3.53)-(3.54) for $f - \Pi_h^r f$, these terms satisfy

$$b(\, f - \Pi_h^r f,\; f - \Pi_h^r f;\, \aleph_h\,) \;+\; \| \, f_0 - \Pi_h^r f(0) \, \|_{0,\Omega}^2 \;\leq\; C\, h^{2\mu - 1}\,. \tag{4.100}$$

We remark that this shows that if the estimate given in (4.98) holds, for the case in which

$\aleph_h$ is taken to be α, then it is optimal in h, since it achieves a convergence order of $2\mu - 1$ in h, which is the same order achieved above by $\Pi_h^r f$.

The remainder of the proof addresses estimating the more formidable term $b(\, F_h - \Pi_h^r f,\, F_h - \Pi_h^r f;\, \aleph_h\,)$. We first note that from (4.59), it follows that

$$\begin{aligned} b\,(\, F_h - \Pi_h^r f, F_h - \Pi_h^r f; \aleph_h\,) \;&=\; \|\, F_h(T) - \Pi_h^r f(T)\,\|_{0,\Omega}^2 \\ &+\; \int_0^T \sum_{k=1}^{P_h} \|\!|\, |\,\overline{\aleph_h}\cdot\nu_k\,|^{1/2}\,[\, F_h - \Pi_h^r f\,]\,|\!\|_{0,\hat{\mathcal{F}}_h}^2 \;+\; \int_0^T \|\!|\, |\,\aleph_h\cdot\nu_k\,|^{1/2}\,(\, F_h - \Pi_h^r f\,)\,|\!\|_{0,\Gamma_O}^2 \\ &+\; \int_0^T \|\!|\, |\,\aleph_h\cdot\nu_k\,|^{1/2}\,(\, F_h - \Pi_h^r f\,)\,|\!\|_{0,\Gamma_I}^2 \,. \end{aligned} \tag{4.101}$$

The estimation problem now proceeds by finding an upper bound for $b(F_h - \Pi_h^r f, F_h - \Pi_h^r f; \aleph_h)$ that depends on the three differences $f - \Pi_h^r f$, $\alpha - \aleph_h$, and $F_h - \Pi_h^r f$. The differences $f - \Pi_h^r f$ and $\alpha - \aleph_h$ will then be estimated using the interpolation properties of $\Pi_h^r f$ and the approximation properties of $\aleph_h$. The remaining expressions, which will only depend on $F_h - \Pi_h^r f$, will be absorbed by the righthandside terms in (4.101) and by Lemma 9.

To begin this process, we use the linearity of $b(\,\cdot\,, F_h - \Pi_h^r f)$ and the pseudo-Galerkin orthogonality relationship (4.94) to get that

$$\begin{aligned} b(\, F_h - \Pi_h^r f,\, F_h - \Pi_h^r f;\, \aleph_h\,) \;&=\; b(\, F_h,\, F_h - \Pi_h^r f;\, \aleph_h\,) \;-\; b(\, \Pi_h^r f,\, F_h - \Pi_h^r f;\, \aleph_h\,) \\ &=\; b(\, f,\, F_h - \Pi_h^r f;\, \alpha\,) \;-\; b(\, \Pi_h^r f,\, F_h - \Pi_h^r f;\, \aleph_h\,) \\ &=\; \big(\; b(\, f,\, F_h - \Pi_h^r f;\, \alpha\,) \;-\; b(\, f,\, F_h - \Pi_h^r f;\, \aleph_h\,)\;\big) \\ &+\; b(\, f - \Pi_h^r f,\, F_h - \Pi_h^r f;\, \aleph_h\,)\,. \end{aligned} \tag{4.102}$$

The task now is to derive suitable bounds for the first two terms and the last term, respectively.

The term $b(f - \Pi_h^r f, F_h - \Pi_h^r f; \aleph_h)$ in (4.102) can be handled using standard techniques, with a few substantial modifications. From the definition of $b\,(\cdot,\cdot)$ given in (4.56), we get

that

$$
\begin{aligned}
b\,(f-\Pi_h^r f, F_h-\Pi_h^r f;\aleph_h) \;&=\; \int_0^T \big(\,(f-\Pi_h^r f)_t, F_h-\Pi_h^r f\,\big)_\Omega \\
&-\; \int_0^T \sum_{j=1}^{N_h} \big(\,\aleph_h(f-\Pi_h^r f), \nabla(F_h-\Pi_h^r f)\,\big)_{K_j} \\
&+\; \int_0^T \sum_{k=1}^{P_h} \big\langle\,(f-\Pi_h^r f^u(\aleph_h))\,\overline{\aleph_h}\,[\,F_h-\Pi_h^r f\,], \nu_k\,\big\rangle_{f_k} \\
&+\; \frac{1}{2}\int_0^T \sum_{k=1}^{P_h} \big\langle\,(f-\overline{\Pi_h^r f})[\,\aleph_h\,]\,[\,F_h-\Pi_h^r f\,], \nu_k\,\big\rangle_{f_k} \\
&+\; \int_0^T \big\langle\,(f-\Pi_h^r f)\aleph_h\cdot\nu_{K_j}, F_h-\Pi_h^r f\,\big\rangle_{\Gamma_O}\,. \\
&=\; T_1\;+\;T_2\;+\;T_3\;+\;T_4\;+\;T_5\,. \qquad (4.103)
\end{aligned}
$$

We now prove estimates for each of these five terms.

T_1 is trivial to estimate. This stems from the fact that $\forall\, t\in[\,0,T\,]$, $(\,f,w\,)_\Omega = (\,\Pi_h^r f, w\,)_\Omega$, $\forall\, w\in D^r(\mathcal{T}_h)$. Since $F_h-\Pi_h^r f\in D^r(\mathcal{T}_h)$, it then follows that

$$
T_1 \;=\; 0\,. \qquad (4.104)
$$

Define $\Pi_h^0\,\alpha$ to be the projection of each component of α into piecewise constants, i.e.,

$$
(\,\alpha\;-\;\Pi_h^0\,\alpha\,,1)_{K_j} \;=\; 0\,, \qquad \forall\; K_j\in\mathcal{T}_h\,. \qquad (4.105)
$$

Then T_2 satisfies

$$
\begin{aligned}
T_2 \;&=\; -\int_0^T \sum_{j=1}^{N_h} (\,\aleph_h\,(f-\Pi_h^r f), \nabla(F_h-\Pi_h^r f)\,)_{K_j} \\
&=\; \int_0^T \sum_{j=1}^{N_h} (\,(\alpha-\aleph_h)(f-\Pi_h^r f), \nabla(F_h-\Pi_h^r f)\,)_{K_j} \\
&-\; \int_0^T \sum_{j=1}^{N_h} (\,(\alpha-\Pi_h^0\,\alpha)(f-\Pi_h^r f), \nabla(F_h-\Pi_h^r f)\,)_{K_j}\,, \qquad (4.106)
\end{aligned}
$$

which follows since each component of $\nabla(F_h-\Pi_h^r f)$ is in $D^r(\mathcal{T}_h)$. Applying the inverse inequality $\|\nabla(F_h-\Pi_h^r f)\,\|_{0,K_j}\le h_j^{-1}\|F_h-\Pi_h^r f\|_{0,K_j}$ to the first term in righthandside of

the above equality leads to

$$
\begin{aligned}
&\int_0^T \sum_{j=1}^{N_h} \big(\, (\alpha - \aleph_h)\,(f - \Pi_h^r f),\, \nabla(F_h - \Pi_h^r f)\,\big)_{K_j} \\
&\leq \int_0^T \sum_{j=1}^{N_h} \| \, (\alpha - \aleph_h)\,(f - \Pi_h^r f)\,\|_{0,K_j} \, h_j^{-1} \| \, F_h - \Pi_h^r f \,\|_{0,K_j} \\
&\leq C\, h^{-1} \int_0^T \| \, (\alpha - \aleph_h)\,(f - \Pi_h^r f)\,\|_{0,\Omega} \, \| \, F_h - \Pi_h^r f \,\|_{0,\Omega} \\
&\leq C\, h^{-1} \sup_{0\leq t\leq T} \| \, f - \Pi_h^r f \,\|_{L^\infty(\Omega)} \int_0^T \| \, \alpha - \aleph_h \,\|_{0,\Omega} \, \| \, F_h - \Pi_h^r f \,\|_{0,\Omega} \\
&\leq C\, h^{s-4} \sup_{0\leq t\leq T} \| \, f - \Pi_h^r f \,\|_{0,\Omega} \int_0^T \| \, \alpha - \aleph_h \,\|_{0,\Omega} \, \| \, F_h - \Pi_h^r f \,\|_{0,\Omega} \\
&\leq C\, h^{\mu+s-4} \int_0^T \| \, \alpha - \aleph_h \,\|_{0,\Omega} \, \| \, F_h - \Pi_h^r f \,\|_{0,\Omega} \, , \qquad (4.107)
\end{aligned}
$$

where the second to the last line follows from that fact that $s > 3$ and inequality (2.104).

The interpolaton inequality $\| \, \alpha - \Pi_h^0 \, \alpha \,\|_{L^\infty(\Omega)} \leq C\, h \, \|\alpha\|_{W^{1,\infty}(\Omega)}$ implies that the second term in the righthandside of (4.106) obeys

$$
\begin{aligned}
&- \int_0^T \sum_{j=1}^{N_h} \big(\, (\alpha - \Pi_h^0 \, \alpha)(f - \Pi_h^r f), \nabla(F_h - \Pi_h^r f)\,\big)_{K_j} \, dt \\
&\leq \| \, \alpha - \Pi_h^0 \, \alpha \,\|_{L^\infty(\Omega)} \int_0^T \sum_{j=1}^{N_h} \| \, f - \Pi_h^r f \,\|_{0,K_j} \, \| \, \nabla(F_h - \Pi_h^r f)\,\|_{0,K_j} \, dt \\
&\leq C \, \| \, \alpha \,\|_{W^{1,\infty}(\Omega)} \int_0^T \left(\sum_{j=1}^{N_h} \| \, f - \Pi_h^r f \,\|_{0,K_j}^2 \right)^{1/2} \left(\sum_{j=1}^{N_h} \| \, F_h - \Pi_h^r f \,\|_{0,K_j}^2 \right)^{1/2} dt \\
&\leq C \int_0^T \| \, f - \Pi_h^r f \,\|_{0,\Omega} \, \| \, F_h - \Pi_h^r f \,\|_{0,\Omega} \\
&\leq C\, h^{\mu} \int_0^T \| \, F_h - \Pi_h^r f \,\|_{0,\Omega} \, . \qquad (4.108)
\end{aligned}
$$

Inserting the bounds from the two inequalities above into (4.106) yields that

$$
T_2 \leq C\, h^{\mu+s-4} \int_0^T \| \, \alpha - \aleph_h \,\|_{0,\Omega} \, \| \, F_h - \Pi_h^r f \,\|_{0,\Omega} \; + \; C\, h^{\mu} \int_0^T \| \, F_h - \Pi_h^r f \,\|_{0,\Omega} \, . \quad (4.109)
$$

As for T_3, we have that

$$
\begin{aligned}
T_3 \;&=\; \int_0^T \sum_{k=1}^{P_h} \langle\, (f - (\Pi_h^r f)^u(\aleph_h))\, \overline{\aleph_h}\, [\, F - \Pi_h^r f\,], \nu_k \,\rangle_{f_k} \\
&\leq\; \int_0^T \| \,|\, \overline{\aleph_h} \cdot \nu_k \,|^{1/2}\, (f - (\Pi_h^r f)^u(\aleph_h))\, \|_{0,\mathring{\mathcal{F}}_h} \| \,|\, \overline{\aleph_h} \cdot \nu_k \,|^{1/2}\, [\, F - \Pi_h^r f\,] \,\|_{0,\mathring{\mathcal{F}}_h} \\
&\leq\; C \int_0^T \| \,|\, \alpha \cdot \nu_k \,|^{1/2}\, (f - (\Pi_h^r f)^u(\aleph_h))\, \|^2_{0,\mathring{\mathcal{F}}_h} \\
&\;+\; C \int_0^T \| \,|\, (\alpha - \overline{\aleph_h}) \cdot \nu_k \,|^{1/2}\, (f - (\Pi_h^r f)^u(\aleph_h))\, \|^2_{0,\mathring{\mathcal{F}}_h} + \frac{1}{4} \int_0^T \| \,|\, \overline{\aleph_h} \cdot \nu_k \,|^{1/2}\, [\, F - \Pi_h^r f\,] \,\|^2_{0,\mathring{\mathcal{F}}_h} \\
&\leq\; C \int_0^T \| \,(f - (\Pi_h^r f)^u(\aleph_h))\, \|^2_{0,\mathring{\mathcal{F}}_h} \;+\; C \int_0^T \| \,|\, (\alpha - \overline{\aleph_h}) \cdot \nu_k \,|^{1/2}\, (f - (\Pi_h^r f)^u(\aleph_h))\, \|^2_{0,\mathring{\mathcal{F}}_h} \\
&\;+\; \frac{1}{4} \int_0^T \| \,|\, \overline{\aleph_h} \cdot \nu_k \,|^{1/2}\, [\, F - \Pi_h^r f\,] \,\|^2_{0,\mathring{\mathcal{F}}_h} \,. \qquad (4.110)
\end{aligned}
$$

To simplify the first term in the above estimate for T_3, we perform the following: from the fact that $f_k = K_1 \cap K_2$, we have that

$$
\begin{aligned}
\| \, f(t) - (\Pi_h^r f)^u(\aleph_h) \,\|^2_{0,\mathring{\mathcal{F}}_h} \;&\leq\; 2 \sum_{k=1}^{P_h} \Big(\| \, f - (\Pi_h^r f)_{|K_1} \,\|^2_{0,f_k} \;+\; \| \, f - (\Pi_h^r f)_{|K_2} \,\|^2_{0,f_k} \Big) \\
&\leq\; C\, h^{-1} \sum_{k=1}^{P_h} \Big(\| \, f - \Pi_h^r f \,\|^2_{0,K_1} \;+\; \| \, f - \Pi_h^r f \,\|^2_{0,K_2} \Big) \\
&\leq\; C\, h^{-1} \| \, f - \Pi_h^r f \,\|^2_{0,\Omega} \;\leq\; C\, h^{2\mu - 1} \,, \qquad (4.111)
\end{aligned}
$$

where the last line follows from the fact that the number of faces a boundary element can intersect is uniformly bounded above.

The second term in the estimate for T_3 satisfies

$$
\begin{aligned}
&\| \,|\, (\alpha - \overline{\aleph_h}) \cdot \nu_k \,|^{1/2}\, (f - (\Pi_h^r f)^u(\aleph_h))\, \|^2_{0,\mathring{\mathcal{F}}_h} \\
&\quad\leq\; \sup_{0 \leq t \leq T} \| \, f - \Pi_h^r f \,\|^2_{L^\infty(\Omega)} \int_0^T \sum_{f_k \in \mathring{\mathcal{F}}_h} \| \,|\, (\alpha - \overline{\aleph_h}) \cdot \nu_k \,|^{1/2} \,\|^2_{0,\mathring{\mathcal{F}}_h} \\
&\quad\leq\; C\, h^{2\mu + 2s - 6} \sum_{f_k \in \mathring{\mathcal{F}}_h} |f_k|^{5/2} \, \| \, \alpha - \overline{\aleph_h} \,\|_{0,f_k} \;\leq\; C\, h^{2\mu + 2s - 6} \Big(\sum_{f_k \in \mathring{\mathcal{F}}_h} h^5 \Big)^{1/2} \| \, \alpha - \overline{\aleph_h} \,\|_{0,\mathring{\mathcal{F}}_h} \\
&\quad\leq\; C\, h^{2\mu + 2s - 7} \| \, \alpha - \overline{\aleph_h} \,\|_{0,\mathring{\mathcal{F}}_h} \;\leq\; C\, h^{2\mu + 2s - 8} \| \, \alpha - \aleph_h \,\|_{0,\Omega} \,, \qquad (4.112)
\end{aligned}
$$

where the second to the last line follows from the fact that the number of faces a boundary element can intersect is uniformly bounded above and the fact that $N_h\, h^6 \sim |\,\Omega\,|$.

Upon plugging the above two estimates into the inequality for T_3, we arrive at

$$\begin{aligned} T_3 \;&\leq\; C\, h^{2\mu-1} \;+\; C\, h^{2\mu+2s-8} \int_0^T \|\!|\, \alpha - \aleph_h \,|\!\|_{0,\Omega} \\ &+\; \frac{1}{4} \int_0^T \|\!|\, |\, \overline{\aleph_h} \cdot \nu_k \,|^{1/2} \,[\, F - \Pi_h^r\, f \,]\, |\!\|^2_{0,\hat{\mathcal{F}}_h} \,. \end{aligned} \tag{4.113}$$

T_4 is seen to satisfy

$$\begin{aligned} T_4 \;&=\; \frac{1}{2} \int_0^T \sum_{k=1}^{P_h} \langle\, (f - \overline{\Pi_h^r\, f})\, [\, \aleph_h - \alpha \,]\, [\, F - \Pi_h^r\, f \,], \nu_k \,\rangle_{f_k} \\ &\leq\; \frac{1}{2} \int_0^T \|\!|\, [\, \aleph_h \,]\, (f - \overline{\Pi_h^r\, f}) \,|\!\|_{0,\hat{\mathcal{F}}_h} \,\|\!|\, [\, F - \Pi_h^r\, f \,]\, |\!\|_{0,\hat{\mathcal{F}}_h} \\ &\leq\; C\, h^{\mu+s-3} \int_0^T \|\!|\, [\, \aleph_h \,] \,|\!\|_{0,\hat{\mathcal{F}}_h} \,\|\!|\, [\, F - \Pi_h^r\, f \,]\, |\!\|_{0,\hat{\mathcal{F}}_h} \,. \end{aligned} \tag{4.114}$$

We now focus our attention on the term in the above integrand involving $F - \Pi_h^r\, f$. First, we write that

$$\begin{aligned} \|\, [\, F_h - \Pi_h^r\, f \,]\, \|_{0,f_k} \;&\leq\; \|\, (F_h - \Pi_h^r\, f)_{|K_1} \,\|_{0,f_k} \;+\; \|\, (F_h - \Pi_h^r\, f)_{|K_2} \,\|_{0,f_k} \\ &\leq\; C\, h^{-1/2} \,\|\, F_h - \Pi_h^r\, f \,\|_{0,K_{12}} \,. \end{aligned} \tag{4.115}$$

This implies

$$\begin{aligned} \|\!|\, [\, F - \Pi_h^r\, f \,]\, |\!\|_{0,\hat{\mathcal{F}}_h} \;&\leq\; C\, h^{-1/2} \left(\sum_{k=1}^{P_h} \|\, F_h - \Pi_h^r\, f \,\|^2_{0,K_{12}} \right)^{1/2} \\ &=\; C\, h^{-1/2} \,\|\!|\, F_h - \Pi_h^r\, f \,|\!\|_{0,\Omega} \,. \end{aligned} \tag{4.116}$$

So, the above results show that T_4 obeys

$$T_4 \;\leq\; C\, h^{\mu+s-7/2} \int_0^T \|\!|\, [\, \aleph_h \,] \,|\!\|_{0,\hat{\mathcal{F}}_h} \,\|\!|\, F_h - \Pi_h^r\, f \,|\!\|_{0,\Omega} \,. \tag{4.117}$$

In the same way that the estimate for T_3 was derived, it follow that T_5 satisfies

$$\begin{aligned} T_5 \;&=\; \int_0^T \sum_{f_k \in \Gamma_O} \langle\, (f - \Pi_h^r f)\, \aleph_h \cdot \nu_k,\, F_h - \Pi_h^r f \,\rangle_{f_k} \\ &\leq\; C\, h^{2\mu-1} \;+\; C\, h^{2\mu+2s-8} \int_0^T \| \alpha - \aleph_h \|_{0,\Omega} \;+\; \frac{1}{4} \int_0^T \| \, |\, \aleph_h \cdot \nu_k \,|^{1/2} \, (F_h - \Pi_h^r f) \, \|^2_{0,\Gamma_O} \,. \end{aligned} \tag{4.118}$$

This completes the estimation of the terms T_1, T_2, T_3, T_4, T_5. Upon combining these estimates, we get that

$$\begin{aligned} & b(\, f - \Pi_h^r f,\, F_h - \Pi_h^r f;\, \aleph_h \,) \;=\; T_1 \;+\; T_2 \;+\; T_3 \;+\; T_4 \;+\; T_5 \\ & \leq\; \frac{1}{4} \int_0^T \| \, |\, \overline{\aleph_h} \cdot \nu_k \,|^{1/2} \, [\, F - \Pi_h^r f \,] \, \|^2_{0,\mathring{\mathcal{F}}_h} \;+\; \frac{1}{4} \int_0^T \| \, |\, \aleph_h \cdot \nu_k \,|^{1/2} \, (F_h - \Pi_h^r f) \, \|^2_{0,\Gamma_O} \\ & \;+\; C\, h^{2\mu-1} \;+\; C\, h^{2\mu+2s-8} \int_0^T \| \alpha - \aleph_h \|_{0,\Omega} \\ & \;+\; C \int_0^T \Big(h^{\mu} \;+\; h^{\mu+s-4} \| \alpha - \aleph_h \|_{0,\Omega} \;+\; h^{\mu+s-7/2} \| \, [\, \aleph_h \,] \, \|_{0,\mathring{\mathcal{F}}_h} \Big) \| F_h - \Pi_h^r f \|_{0,\Omega} \,. \end{aligned} \tag{4.119}$$

We now would like to simplify the above righthandside estimate further, by exploiting the structure of α and $\aleph_h$. Since $\alpha - \aleph_h = (0, \nabla_x(\psi - \psi_h))$ we see that

$$\| \alpha - \aleph_h \|_{0,\Omega} \;=\; |\, \Omega^v \,|^{1/2} \, \| \nabla_x(\psi - \psi_h) \|_{0,\Omega^x} \;\leq\; C \, \| \nabla_x(\psi - \psi_h) \|_{0,\Omega^x} \,. \tag{4.120}$$

We now turn to simplifying the term $\| \, [\, \aleph_h \,] \, \|_{0,\mathring{\mathcal{F}}_h}$. To do this, we make use of the partition $\mathcal{F}_h = \mathcal{F}_h^x \cup \mathcal{F}_h^v$. Let us first consider an arbitrary face $f_k \in \mathring{\mathcal{F}}_h^v$. Let $K_{j_x}^x \in \mathcal{T}_h^x$ and $f_{k_v}^v \in \mathring{\mathcal{G}}_h^v$ be such that $f_k = K_{j_x}^x \cup f_{k_v}^v$. Since $\nabla_x \psi_h$ is continuous in $K_{j_x}^x$ and since v is continuous on Ω^v, it follows that

$$[\, \aleph_h \,] \;\equiv\; 0 \,, \quad \text{on } f_k \,.$$

Thus,

$$\| \, [\, \aleph_h \,] \, \|_{0,\mathring{\mathcal{F}}_h} \;=\; \| \, [\, \aleph_h \,] \, \|_{0,\mathring{\mathcal{F}}_h^x} \,.$$

We continue simplifying by noting that

$$\begin{aligned}
\| [\,\aleph_h\,] \|^2_{0,\mathring{\mathcal{F}}_h^x} \;&=\; \sum_{f_{k_x}^x \in \mathring{\mathcal{G}}_h^x} \sum_{\substack{j_v=1 \\ f_k = f_{k_x}^x \cup K_{j_v}^v}}^{N_h^v} \| [\,\aleph_h\,] \|^2_{0,f_k} \;=\; \sum_{f_{k_x}^x \in \mathring{\mathcal{G}}_h^x} \sum_{j_v=1}^{N_h^v} h_v^3 \, \| [\,\nabla_x \psi_h\,] \|^2_{0,f_{k_x}^x} \\
&\leq\; C \, \| [\,\nabla_x \psi_h\,] \|^2_{0,\mathring{\mathcal{G}}_h^x} \,, \qquad (4.121)
\end{aligned}$$

since $\nabla_x \psi_h$ is independent of v and $N_h^v \, h_v^3 \sim |\Omega^v|$. Hence

$$\| [\,\aleph_h\,] \|_{0,\mathring{\mathcal{F}}_h} \;\leq\; C \, \| [\,\nabla_x \psi_h\,] \|_{0,\mathring{\mathcal{G}}_h^x} \,. \qquad (4.122)$$

Using the above simplifications, it follows that (4.119) implies

$$\begin{aligned}
&b(\, f - \Pi_h^r f,\, F_h - \Pi_h^r f;\, \aleph_h\,) \;=\; T_1 \;+\; T_2 \;+\; T_3 \;+\; T_4 \;+\; T_5 \\
&\leq\; \frac{1}{4} \int_0^T \| \, | \, \overline{\aleph_h} \cdot \nu_k \, |^{1/2} \, [\, F - \Pi_h^r f \,] \, \|^2_{0,\mathring{\mathcal{F}}_h} \;+\; \frac{1}{4} \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, (F_h - \Pi_h^r f) \, \|^2_{0,\Gamma_O} \\
&+\; C \, h^{2\mu-1} \;+\; C \, h^{2\mu+2s-8} \int_0^T \| \, \nabla_x (\psi - \psi_h) \, \|_{0,\Omega^x} \\
&+\; C \int_0^T \Big(\, h^{\mu} \;+\; h^{\mu+s-4} \, \| \, \nabla_x (\psi - \psi_h) \, \|_{0,\Omega^x} \;+\; h^{\mu+s-7/2} \| \, [\,\nabla_x \psi_h\,] \, \|_{0,\mathring{\mathcal{G}}_h^x} \Big) \, \| \, F_h - \Pi_h^r f \, \|_{0,\Omega} \,. \qquad (4.123)
\end{aligned}$$

We now focus our attention on bounding the first two terms in (4.102). We note the following estimates will utilize that $\sup_{0 \leq t \leq T} \| \, f(t) \, \|_{L^\infty(\Omega)} < \infty$, which holds by the assumed regularity of f. By the definition of $b\,(\cdot,\cdot)$ given in (4.56) we have that

$$\begin{aligned}
&b(\, f,\, F_h - \Pi_h^r f;\, \alpha\,) \;-\; b(\, f,\, F_h - \Pi_h^r f;\, \aleph_h\,) \\
&=\; -\int_0^T \sum_{j=1}^{N_h} (\, (\alpha - \aleph_h)\, f, \nabla (F_h - \Pi_h^r f)\,)_{K_j} \;+\; \int_0^T \sum_{k=1}^{P_h} \big\langle \, f \, (\, \alpha - \overline{\aleph_h}\,) \, [\, F_h - \Pi_h^r f \,],\, \nu_k \, \big\rangle_{f_k} \\
&+\; \frac{1}{2} \int_0^T \sum_{k=1}^{P_h} \big\langle \, f \, [\, \alpha - \aleph_h \,] \, [\, F_h - \Pi_h^r f \,],\, \nu_k \, \big\rangle_{f_k} \;+\; \int_0^T \sum_{f_k \in \Gamma_O} \big\langle \, f \, (\alpha - \aleph_h) \cdot \nu_k,\, F_h - \Pi_h^r f \, \big\rangle_{f_k} \\
&=\; D_1 \;+\; D_2 \;+\; D_3 \;+\; D_4 \,. \qquad (4.124)
\end{aligned}$$

We note that the above equality could be simplified since $[\,\alpha\,] = 0$ on the interior faces of the mesh. However, to maintain clarity we choose to live as it is written. The task now is

to obtain reasonable estimates for each of the four terms in the righthandside of (4.124).

We begin by first rewriting D_1 in the following manner: For D_1, the fact that α and $\aleph_h$ are both divergence free implies

$$\begin{aligned} D_1 &= -\int_0^T \sum_{j=1}^{N_h} \big((\alpha - \aleph_h)\, f, \nabla(F_h - \Pi_h^r f) \big)_{K_j} \\ &= -\int_0^T \sum_{j=1}^{N_h} \big\langle (\alpha - \aleph_h)\ (F_h - \Pi_h^r f),\ f\, \nu_{K_j} \big\rangle_{\partial K_j} \\ &= -\int_0^T \sum_{f_k \in \mathring{\mathcal{F}}_h} \big\langle [\,(\alpha - \aleph_h)\,(F_h - \Pi_h^r f)\,],\ f\, \nu_k \big\rangle_{f_k} \\ &\quad - \int_0^T \sum_{f_k \in \Gamma} \big\langle (\alpha - \aleph_h)\,(F_h - \Pi_h^r f),\ f\, \nu_k \big\rangle_{f_k} . \end{aligned} \tag{4.125}$$

Substituting the above righthandside for D_1 into (4.124) yields the equation

$$\begin{aligned} D_1 \ + \ D_2 \ + \ D_3 \ + \ D_4 \ &= \ \int_0^T \sum_{k=1}^{P_h} \big\langle \ -[\,(\alpha - \aleph_h)\,(F_h - \Pi_h^r f)\,] \\ &+ \ (\,\alpha - \overline{\aleph_h}\,)\,[\,F_h - \Pi_h^r f\,] \ + \ \frac{1}{2}\,[\,\alpha - \aleph_h\,]\,[\,F_h - \Pi_h^r f\,],\, f\, \nu_k \big\rangle_{f_k} \\ &- \int_0^T \sum_{f_k \in \Gamma_I} \big\langle (\alpha - \aleph_h)\,(F_h - \Pi_h^r f),\ f\, \nu_k \big\rangle_{f_k} . \end{aligned} \tag{4.126}$$

We now focus on simplifying the three interior face terms in the above equation.

On a given an interior face f_k, define the quantities $a = \alpha,\ b = \aleph_h,\ c = F_h$, and $d = \Pi_h^r f$. Then let $b_m = (\aleph_h)_{|K_m},\ c_m = (F_h)_{|K_m}$, and $c_m = (F_h)_{|K_m}$ for $m = 1, 2$. We do not bother to define a_m, since $[\,a\,] = [\,\alpha\,] = 0$. Using these quantities, we now perform the following

manipulations:

$$
\begin{aligned}
&-[\,(\alpha-\aleph_h)\,(F_h-\Pi_h^r f)\,]\;+\;(\,\alpha-\overline{\aleph_h}\,)\,[\,F_h-\Pi_h^r f\,]\;+\;\frac{1}{2}\,[\,\alpha-\aleph_h\,]\,[\,F_h-\Pi_h^r f\,]\\
&\quad=\;-[\,(a-b)\,(c-d)\,]\;+\;(\,a-\bar b\,)\,[\,c-d\,]\;+\;\frac{1}{2}\,[\,a-b\,]\,[\,c-d\,]\\
&\quad=\;[\,b\,(c-d)\,]\;-\;\bar b\,[\,c-d\,]\;-\;\frac{1}{2}\,[\,b\,]\,[\,c-d\,]\\
&\quad=\;[\,b\,(c-d)\,]\;-\;(\,\bar b\;+\;\frac{1}{2}\,[\,b\,]\,)\,[\,c-d\,]\\
&\quad=\;[\,b\,(c-d)\,]\;-\;b_1\,[\,c-d\,]\\
&\quad=\;b_1\,(c_1-d_1)\;-\;b_2\,(c_2-d_2)\;-\;b_1\,(\,(c_1-d_1)-(c_2-d_2)\,)\\
&\quad=\;-b_2\,(c_2-d_2)\;+\;b_1\,(c_2-d_2)\\
&\quad=\;[\,b\,]\,(c_2-d_2)\\
&\quad=\;[\,\aleph_h\,]\,(\,(F_h)_{|K_2}\;-\;(\Pi_h^r f)_{|K_2}\,)\,.
\end{aligned}
\tag{4.127}
$$

Using the above identity, we see that

$$
\begin{aligned}
&\int_0^T\sum_{k=1}^{P_h}\langle\;-[\,(\alpha-\aleph_h)\,(F_h-\Pi_h^r f)\,]\;+\;(\,\alpha-\overline{\aleph_h}\,)\,[\,F_h-\Pi_h^r f\,]\\
&\;+\;\frac{1}{2}\,[\,\alpha-\aleph_h\,]\,[\,F_h-\Pi_h^r f\,],\,f\,\nu_k\,\rangle_{f_k}\\
&=\;\int_0^T\sum_{f_k\in\mathring{\mathcal{F}}_h}\langle\,[\,\aleph_h\,]\,(\,(F_h)_{|K_2}\;-\;(\Pi_h^r f)_{|K_2}\,),\,f\,\nu_k\,\rangle_{f_k}\\
&\leq\;C\int_0^T\int_0^T\|\!|\,[\,\aleph_h\,]\,|\!\|_{0,\mathring{\mathcal{F}}_h}\,h^{-1/2}\left(\sum_{f_k\in\mathring{\mathcal{F}}_h}\|\,F_h-\Pi_h^r f\,\|_{0,K_{12}}^2\right)^{1/2}\\
&\leq\;C\,h^{-1/2}\int_0^T\|\!|\,[\,\aleph_h\,]\,|\!\|_{0,\mathring{\mathcal{F}}_h}\,|\!|\!|\,F_h-\Pi_h^r f\,|\!|\!|_{0,\Omega}\\
&\leq\;C\,h^{-1/2}\int_0^T\|\!|\,[\,\nabla_x\psi_h\,]\,|\!\|_{0,\mathring{\mathcal{G}}_h^x}\,|\!|\!|\,F_h-\Pi_h^r f\,|\!|\!|_{0,\Omega}\,,
\end{aligned}
\tag{4.128}
$$

where the last line follows from the identity (4.122).

Therefore, we have that

$$D_1 \;+\; D_2 \;+\; D_3 \;+\; D_4 \;\leq\; C\,h^{-1/2} \int_0^T \|\!|\, [\,\nabla_x \psi_h\,]\,|\!\|_{0,\mathring{\mathcal{G}}_h^x} \,\|\!|\, F_h - \Pi_h^r\, f\,|\!\|_{0,\Omega} \\ - \int_0^T \sum_{f_k \in \Gamma_I} \langle\, (\alpha - \aleph_h)\,(F_h - \Pi_h^r\, f),\; f\,\nu_k\,\rangle_{f_k}\,. \tag{4.129}$$

What now remains is to bound the inflow term in the above righthandside. This is accomplished in part by making use of the fact that $f = 0,\ \forall\, v \in \partial\Omega^v$. We proceed as follows:

$$\int_0^T \sum_{f_k \in \Gamma_I} \langle\, (\alpha - \aleph_h)\,(F_h - \Pi_h^r\, f),\; f\,\nu_k\,\rangle_{f_k} \;\leq\; \int_0^T \|\!|\, (\alpha - \aleph_h)\, f\,|\!\|_{0,\Gamma_I} \,\|\!|\, F_h - \Pi_h^r\, f\,|\!\|_{0,\Gamma_I} \\ \leq\; C\,h^{-1/2} \int_0^T \|\!|\, (\alpha - \aleph_h)\, f\,|\!\|_{0,\Gamma_I} \,\|\!|\, F_h - \Pi_h^r\, f\,|\!\|_{0,\Omega} \\ \leq\; C\,h^{-1/2} \int_0^T \|\!|\, \alpha - \aleph_h\,|\!\|_{0,\mathcal{F}_h^x/\mathring{\mathcal{F}}_h^x} \,\|\!|\, F_h - \Pi_h^r\, f\,|\!\|_{0,\Omega}\,. \tag{4.130}$$

To simplify the above expression, we note that

$$\|\!|\, \alpha - \aleph_h\,|\!\|^2_{0,\mathcal{F}_h^x/\mathring{\mathcal{F}}_h^x} \;=\; \|\!|\, \nabla_x(\psi - \psi_h)\,|\!\|^2_{0,\mathcal{F}_h^x/\mathring{\mathcal{F}}_h^x} \;=\; \sum_{f_{k_x}^x \in \partial\Omega^x} \sum_{\substack{j_v=1 \\ f_k = f_{k_x}^x \cup K_{j_v}^v}}^{N_h^v} |\!|\, \nabla_x(\psi - \psi_h)\,|\!|^2_{f_k} \\ \leq\; |\,\Omega^v\,|\,\|\!|\, \nabla_x(\psi - \psi_h)\,|\!\|^2_{\partial\Omega^x}\,. \tag{4.131}$$

Thus, it follows that

$$\int_0^T \sum_{f_k \in \Gamma_I} \langle\, (\alpha - \aleph_h)\,(F_h - \Pi_h^r\, f),\; f\,\nu_k\,\rangle_{f_k} \\ \leq\; C\,h^{-1/2} \int_0^T \|\!|\, \nabla_x(\psi - \psi_h)\,|\!\|_{\partial\Omega^x} \,\|\!|\, F_h - \Pi_h^r\, f\,|\!\|_{0,\Omega}\,. \tag{4.132}$$

Upon combining all of the above results, we get that

$$b(\, f,\, F_h - \Pi_h^r\, f;\, \alpha\,) \;-\; b(\, f,\, F_h - \Pi_h^r\, f;\, \aleph_h\,) \;=\; D_1 \;+\; D_2 \;+\; D_3 \;+\; D_4 \\ \leq\; C \int_0^T \Big(h^{-1/2}\,\|\!|\, [\,\nabla_x \psi_h\,]\,|\!\|_{0,\mathring{\mathcal{G}}_h^x} \;+\; h^{-1/2}\,\|\!|\, \nabla_x(\psi - \psi_h)\,|\!\|_{0,\partial\Omega^x} \Big)\,\|\!|\, F_h - \Pi_h^r\, f\,|\!\|_{0,\Omega}\,. \tag{4.133}$$

Upon plugging in the final estimates for $T_1, \ldots, T_5$ and $D_1, \ldots, D_4$ into the righthandside of (4.102) and then combining the resulting inequality with (4.101), we get that, after a few algebraic manipulations,

$$
\begin{aligned}
&\| F_h(T) - \Pi_h^r f(T) \|_{0,\Omega}^2 \;+\; \int_0^T \| \, | \, \overline{\aleph_h} \cdot \nu_k \, |^{1/2} \, [\, F_h - \Pi_h^r f \,] \, \|_{0,\mathring{\mathcal{F}}_h}^2 \\
&+ \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, (\, F_h - \Pi_h^r f \,) \, \|_{0,\Gamma_O}^2 \;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, (\, F_h - \Pi_h^r f \,) \, \|_{0,\Gamma_I}^2 \\
&\leq \; \frac{1}{4} \int_0^T \| \, | \, \overline{\aleph_h} \cdot \nu_k \, |^{1/2} \, [\, F - \Pi_h^r f \,] \, \|_{0,\mathring{\mathcal{F}}_h}^2 \;+\; \frac{1}{4} \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, (F_h - \Pi_h^r f) \, \|_{0,\Gamma_O}^2 \\
&+ \; C \, h^{2\mu-1} \;+\; C \, h^{2\mu+2s-8} \int_0^T \| \, \nabla_x (\psi - \psi_h) \, \|_{0,\Omega^x} \\
&+ \; C \int_0^T (\, h^{\mu} \;+\; h^{\mu+s-4} \| \, \nabla_x (\psi - \psi_h) \, \|_{0,\Omega^x} \;+\; h^{\mu+s-7/2} \| \, [\, \nabla_x \psi_h \,] \, \|_{0,\mathring{\mathcal{G}}_h^x} \\
&+ \; h^{-1/2} \| \, [\, \nabla_x \psi_h \,] \, \|_{0,\mathring{\mathcal{G}}_h^x} \;+\; h^{-1/2} \| \, \nabla_x (\psi - \psi_h) \, \|_{0,\partial\Omega^x} \Big) \| \, F_h - \Pi_h^r f \, \|_{0,\Omega} \, . \qquad (4.134)
\end{aligned}
$$

In order to simplify inequality (4.101), define the functions $R, A, B : [\, 0, T \,] \ni t \to \mathbb{R}$ in the following way:

$$
\begin{aligned}
R(t) \;=\; &\int_0^T \| \, | \, \overline{\aleph_h} \cdot \nu_k \, |^{1/2} \, [\, F_h - \Pi_h^r f \,] \, \|_{0,\mathring{\mathcal{F}}_h}^2 \;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, (\, F_h - \Pi_h^r f \,) \, \|_{0,\Gamma_O}^2 \\
&+ \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, (\, F_h - \Pi_h^r f \,) \, \|_{0,\Gamma_I}^2 \, , \qquad (4.135)
\end{aligned}
$$

$$
A(t) \;=\; C \left(h^{2\mu-1} \;+\; C \, h^{2\mu+2s-8} \int_0^T \| \, \nabla_x (\psi - \psi_h) \, \|_{0,\Omega^x} \right) , \qquad (4.136)
$$

and

$$
\begin{aligned}
B(t) \;=\; C \Big(&h^{\mu} \;+\; h^{\mu+s-4} \| \, \nabla_x (\psi - \psi_h) \, \|_{0,\Omega^x} \;+\; h^{\mu+s-7/2} \| \, [\, \nabla_x \psi_h \,] \, \|_{0,\mathring{\mathcal{G}}_h^x} \\
&+ \; h^{-1/2} \| \, [\, \nabla_x \psi_h \,] \, \|_{0,\mathring{\mathcal{G}}_h^x} \;+\; h^{-1/2} \| \, \nabla_x (\psi - \psi_h) \, \|_{0,\partial\Omega^x} \Big) . \qquad (4.137)
\end{aligned}
$$

We note that $b(F_h - \Pi_h^r f, F_h - \Pi_h^r f; \aleph_h) = \| F_h(T) - \Pi_h^r f(T) \|_{0,\Omega}^2 + R(T)$, which is the original quantity that we wanted to bound. Using these definitions, inequality (4.101) can

be rewritten, after some algebraic manipulation, as

$$\| F_h(T) - \Pi_h^r f(T) \|_{0,\Omega}^2 \;+\; R(T) \;\leq\; A(T) \;+\; \int_0^T B(t) \,|\!|\!| F_h(t) - \Pi_h^r f(t) |\!|\!|_{0,\Omega} \, dt \,. \quad (4.138)$$

The above inequality is written in a form such that Lemma 9 can be applied to it. Applying this lemma leads to the estimate

$$\begin{aligned}
& b(\, F_h - \Pi_h^r f,\, F_h - \Pi_h^r f;\, \aleph_h \,) \;\leq\; \left(\sup_{0\leq t\leq T} A^{1/2}(t) \;+\; \int_0^T B(t)\, dt \right)^2 \\
& \leq\; 2 \left(\sup_{0\leq t\leq T} A(t) \;+\; \left(\int_0^T B(t)\, dt \right)^2 \right) \\
& \leq\; 2 \left(\sup_{0\leq t\leq T} A(t) \;+\; T \int_0^T B^2(t)\, dt \right) \\
& \leq\; C \left(h^{2\mu-1} \;+\; h^{2\mu+2s-8} \int_0^T |\!|\!| \nabla_x(\psi - \psi_h) |\!|\!|_{0,\Omega^x} \right) \\
& \;+\; C \int_0^T \left(h^{2\mu+2s-8} \,|\!|\!| \nabla_x(\psi - \psi_h) |\!|\!|_{0,\Omega^x}^2 \;+\; h^{2\mu+2s-7} |\!|\!| [\, \nabla_x \psi_h \,] |\!|\!|_{0,\mathring{\mathcal{G}}_h^x}^2 \right. \\
& \;+\; \left. h^{-1} \,|\!|\!| [\, \nabla_x \psi_h \,] |\!|\!|_{0,\mathring{\mathcal{G}}_h^x}^2 \;+\; h^{-1} \,|\!|\!| \nabla_x(\psi - \psi_h) |\!|\!|_{0,\partial\Omega^x}^2 \right), \qquad (4.139)
\end{aligned}$$

where the last line results after applying a few repeated applications of Young's inequality. It is important to note the constant C now depends linearly on T, which is a direct result of using Lemma 9.

By inserting (4.139) and (4.100) into (4.99) and then using identity (4.96), we obtain the desired estimate (4.97) □

The estimate (4.97) in the above theorem is not unique. There are other quantities that depend on the difference of ψ and ψ_h that could have been derived for use in the righthand-side of (4.97). However, the estimate given by Theorem 14 has been derived in light of the fact that our ultimate goal in this work is to propose a DG method to approximate the Vlasov-Poisson system such that an error estimate can be established between a discrete solution resulting from this method and the classical solution of the Vlasov-Poisson system. What is then crucial in the above theorem is that it is structured in such a way that its result can be used in conjunction with Gronwall's lemma or Lemma 9, or both together, to establish the desired error result for the Vlasov-Poisson system. We will soon see that this is indeed the case for the above theorem.

4.5.3 Extension of *a priori* error analysis for controlled flow perturbations

We now investigate the exact nature of the estimate in Theorem 14 for the case when there exists a sequence of local polynomial potential perturbations $\{ \nabla_x \psi_h \}_{h>0} \subset D^r(\mathcal{T}_h)$ defined on the family of meshes $\{ \mathcal{T}_h^x \}_{h>0}$, in which a known approximation result holds for the difference $\nabla_x(\psi - \psi_h)$.

The exact approximation property that will be assumed to hold is that for the data trio (α, f_0, f_I), with the additional regularity condition that the potential $\psi(t) \in H^{\bar{s}}(\mathcal{T}_h^x),\ \forall\, t \in [0,T]$, for some $\bar{s} > 3/2$, there exists a mesh independent constant C such that the following inequality

$$\| \nabla_x(\psi - \psi_h) \|_{0,K^x_{j_x}} \ \leq\ C\, \frac{h^{\bar{\mu}-1}}{r^{\bar{s}-1}}\, \| \psi \|_{\bar{s},K^x_{j_x}}\ , \tag{4.140}$$

holds $K^x_{j_x} \in \mathcal{T}_h^x, \forall\, h > 0$, where $\bar{\mu} = \min\{\bar{r}+1, \bar{s}\}$.

Corollary 2. *Let (α, f_0, f_I) be a compatible trio, with the additional condition that $\psi(t) \in H^{\bar{s}}(\mathcal{T}_h^x),\ \forall\, t \in [0,T]$, for some $\bar{s} > 3/2$. Assume that there exists a unique classical solution f to the Vlasov-Poisson system defined by (α, f_0, f_I), where $f \in C^1([0,T], H^s(\mathcal{T}_h))$, $s > 3$, is satisfied. Let $\{ (\aleph_h, f_0, f_I) \}_{h>0}$ be a sequence of perturbed compatible trios, where the perturbed potentials $\{ \psi_h \}_{h>0}$ satisfy $\psi_h \in D^r(\mathcal{T}_h^x)$ and the estimate*

$$\| \nabla_x(\psi - \psi_h) \|_{0,K^x_{j_x}} \ \leq\ C\, \frac{h^{\bar{\mu}-1}}{r^{\bar{s}-1}}\, \| \psi \|_{\bar{s},K^x_{j_x}}\ , \qquad \forall\ K^x_{j_x} \in \mathcal{T}_h^x,\ \forall\ h > 0\,, \tag{4.141}$$

where $\bar{\mu} = \min\{ \bar{r}+1,\ \bar{s} \}$. Moreover, further assume that

$$\bar{\mu}\ \geq\ \begin{cases} \max\{\ 8 - 2s,\ \mu + 3/2 \} & , \quad \text{if}\ \ 3 < s \leq 7/2\,, \\ \max\{ 9/2 - s,\ \mu + 3/2 \} & , \quad \text{if}\ \ s > 7/2\,, \end{cases} \tag{4.142}$$

where $\mu = \min\{ r+1,\ s \}$. For each $h > 0$, let $F_h \in D^r(\mathcal{T}_h)$ be the discrete DFUG solution to the perturbed Vlasov system defined by $(\aleph_h, f_0, f_I)$. Then the following a priori error

estimate holds:

$$\begin{aligned}
&\| f(T) - F_h(T) \|_{0,\Omega}^2 \;+\; \int_0^T \sum_{k=1}^{P_h} \| \, | \, \overline{\aleph_h} \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,f_k}^2 \\
&\;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,\Gamma_O}^2 \;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,\Gamma_I}^2 \\
&\leq\; C \, h^{2\mu-1} \left(1 \;+\; \frac{1}{\bar{r}^{2\bar{s}-3}} \right) . \qquad (4.143)
\end{aligned}$$

Proof. From the given assumptions, it follows that Theorem 14 can be applied to the estimate (4.97). So, to complete the proof, each of the terms in (4.97) that depend on the difference between ψ and ψ_h must be estimated using the bound (4.141). Using this bound, it is straightforward to establish the following estimates:

$$\begin{aligned}
\| \! | \, [\, \nabla_x \psi_h \,] \, \| \! |_{0,\mathring{\mathcal{G}}_h^x} \;&\leq\; C \, \frac{h^{\bar{\mu}-3/2}}{\bar{r}^{\bar{s}-3/2}} \qquad \text{and} \\
\| \! | \nabla_x (\psi - \psi_h) \| \! |_{0,\partial\Omega^x} \;&\leq\; C \, \frac{h^{\bar{\mu}-3/2}}{\bar{r}^{\bar{s}-3/2}} \, . \qquad (4.144)
\end{aligned}$$

Upon plugging in the above results and the bound (4.141) into the righthandside estimate given in (4.97), we get that, after integrating the time variable,

$$\begin{aligned}
&\| f(T) - F_h(T) \|_{0,\Omega}^2 \;+\; \int_0^T \sum_{k=1}^{P_h} \| \, | \, \overline{\aleph_h} \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,f_k}^2 \\
&\;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,\Gamma_O}^2 \;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,\Gamma_I}^2 \\
&\leq\; C \left(h^{2\mu-1} \;+\; h^{2\mu+2s-8} \, \frac{h^{\bar{\mu}-1}}{r^{\bar{s}-1}} \;+\; h^{2\mu+2s-8} \, \frac{h^{2\bar{\mu}-2}}{r^{2\bar{s}-2}} \;+\; h^{2\mu+2s-7} \frac{h^{2\bar{\mu}-3}}{r^{2\bar{s}-3}} \;+\; h^{-1} \, \frac{h^{2\bar{\mu}-3}}{r^{2\bar{s}-3}} \right) \\
&\leq\; C \left(h^{2\mu-1} \;+\; h^{2\mu+2s-8} \, \frac{h^{\bar{\mu}-1}}{r^{\bar{s}-1}} \;+\; h^{2\mu+2s-7} \frac{h^{2\bar{\mu}-3}}{r^{2\bar{s}-3}} \;+\; h^{-1} \, \frac{h^{2\bar{\mu}-3}}{r^{2\bar{s}-3}} \right) \\
&=\; C \left(h^{2\mu-1} \;+\; \frac{h^{2\mu+2s+\bar{\mu}-9}}{r^{2\bar{s}-2}} \;+\; \frac{h^{2\mu+2s+2\bar{\mu}-10}}{r^{2\bar{s}-3}} \;+\; \frac{h^{2\bar{\mu}-4}}{r^{2\bar{s}-3}} \right) , \qquad (4.145)
\end{aligned}$$

where the second to the last line follows since $h/r \leq 1$.

The righthandside bound in (4.145) will only achieve the optimal order of $2\mu - 1$ in h if $\bar{\mu}$

satisfies the following three condtions:

$$(i) \qquad 2s \;+\; \bar{\mu} \;-\; 9 \geq -1 \qquad \text{and} \tag{4.146}$$

$$(ii) \qquad 2s \;+\; 2\bar{\mu} \;-\; 10 \geq -1 \qquad \text{and} \tag{4.147}$$

$$(iii) \qquad 2\bar{\mu} \;-\; 4 \;\geq 2\mu \;-\; 1\,. \tag{4.148}$$

Conditions (4.146)-(4.148) will be satisfied if and only if

$$\bar{\mu} \geq \begin{cases} \max\{\;8 - 2s,\; \mu + 3/2\;\} & , \quad \text{if } \; 3 \;<\; s \;\leq\; 7/2\,, \\ \max\{\,9/2 - s,\; \mu + 3/2\,\} & , \quad \text{if } \; s \;>\; 7/2\,. \end{cases} \tag{4.149}$$

Since $\bar{\mu}$ is assumed to satisfy these conditions, it then follows that the estimate (4.145) reduces to (4.143). □

Remark 4. *Corollary 2 demonstrates that the discrete DFUG solution to a perturbed Vlasov-System is only guaranteed to be an optimal approximation to the classical solution of the Vlasov-Poisson system of interest, if the gradient of the perturbed potential defining the perturbed system is an approximation to $\nabla_x \psi$ with a high degree of accuracy. In particular, if the solution f and the potential ψ are both smooth enough, then the condtion (4.142) on $\bar{\mu}$ reduces to the condition that*

$$\bar{r} \;\geq\; r \;+\; \frac{3}{2}\,, \tag{4.150}$$

which is equivalent to $\bar{r} \geq r + 2$, since $\bar{r}$ and r are both integers. This appears reasonable, since the DFUG formulation, and the UG formulation as well, use the average values of the flow at the interior faces to determine the "upwind" direction. So, if the perturbed flow does not approximate the true flow with a high degree of accuracy on the faces of the mesh, then the resulting "upwind" direction determined by the perturbed flow could in fact be chosen to be the wrong direction. This has the potential to lead to significant errors in the approximate solution, especially when the true solution has sharp gradients, since at any given interior face the values of the solution might be drastically different on the two elements whose intersection is the face.

4.6 DFUG-NISPG approximation to the Vlasov-Poisson system

In this section, we propose a DG method for approximating the Vlasov-Poisson system. The proposed method combines the NIPG method for the approximation of the Poisson system with the DFUG method for the approximation of the Vlasov system. Hence, the name given to this method is the DFUG-NIPG method. The most significant result of this section is that an *a priori* error estimate will be proved for DFUG-NIPG approximation to the Vlasov-Poisson system when there exists a discrete solution of the DFUG-NIPG formulation. The existence of such a discrete solution will be subject of future work to come. The goal here is propose a method for the Vlasov-Poisson system in which error estimates can be established.

The analysis that was done for the perturbed Poisson and Vlasov systems will play a direct role in obtaining the forthcoming results. The main result of this section is that an *a priori* error estimate is proved between the true solution of the Vlasov-Poisson system and the discrete DFUG-NIPG approximation of this solution.

4.6.1 Weak problem statement

We now state the definitions of a discrete solution to the DFUG-NIPG formulation of the Vlasov-Poisson system (4.20)-(4.25). These definitions follow directly from the definitions of the discrete DFUG solution to the Vlasov system and the discrete NIPG solution to the Poisson system. To simplify the notation in the problem statement, we first define a time-dependent bilinear operator B, which is in fact the bilinear operator that defines the DFUG method. The definition given for B is as follows: $\forall\, (\,\xi, w\,) \in C^1(\,[\,0,T\,],\, H^1(\mathcal{T}_h)\,)\times H^1(\mathcal{T}_h)\,,$

$$B(\,\xi, w;\,\alpha\,) \;=\; (\,\xi_t, w\,)_\Omega \;-\; \sum_{j=1}^{N_h} (\,\alpha\,\xi, \nabla w\,)_{K_j} \;+\; \sum_{k=1}^{P_h} \langle\, \xi^u(\alpha)\,\overline{\alpha}\,[\,w\,] \;+\; \frac{1}{2}\,\overline{\xi}\,[\,\alpha\,]\,[\,w\,], \nu_k\,\rangle_{f_k}$$
$$+ \sum_{f_k\in\Gamma_O} \langle\, \xi\,\alpha\cdot\nu_k, w\,\rangle_{f_k}\,. \tag{4.151}$$

With B defined, the definition of the discrete DFUG-NIPG solution to the Vlasov-Poisson is easily stated.

Definition 17. [Discrete DFUG − NIPG Solution to Vlasov − Poisson System] *Given*

$T > 0$ *and a compatible trio* (f_0, f_I, r_D), *a function pair*

$$(F_h,\, \psi_h) \in \Big(\big(C^1([\,0,T\,],\, D^r(\mathcal{T}_h) \big),\ \big(C^1([\,0,T\,],\, D^{\bar{r}}(\mathcal{T}_h^x) \big) \Big)\ ,$$

r, $\bar{r} \in \mathbb{N}$, *is said to be a discrete DFUG-NIPG solution to the Vlasov-Poisson system (4.20)-(4.25) if the equations*

$$(i) \qquad (F_h(\,t=0\,),\, w\,)_{K_j} \ = \ (f_0,\, w\,)_{K_j}\ , \qquad \textit{and} \tag{4.152}$$

$$(ii) \qquad \forall\ t \in (\,0,T\,],$$

$$\begin{aligned} & B(\,F_h, w;\, \aleph_h\,) \ + \ a(\,\psi_h, \theta\,) \ + \ J(\,\psi_h, \theta\,) \ - \ (\,\rho(F_h),\, \theta\,)_\Omega \\ & = \ - \sum_{f_k \in \Gamma_I} \langle\, f_I\, \aleph_h \cdot \nu_k, w \,\rangle_{f_k} \ + \sum_{f^x_{k_x} \in \Gamma^x_D} \langle\, \nabla_x \theta \cdot \nu^x_{k_x},\, r_D(x,t)\ \rangle_{f^x_{k_x}} \\ & \quad + \sum_{f^x_{k_x} \in \Gamma^x_D} \frac{\bar{r}\sigma_k}{|\, f^x_{k_x}\,|^{1/2}} \langle\, r_D(x,t),\, \theta\,\rangle_{f^x_{k_x}} \end{aligned} \tag{4.153}$$

are satisfied, $\forall$ *test function pairs* $(\,w,\,\theta\,) \in D^r(\mathcal{T}_h) \times D^{\bar{r}}(\mathcal{T}_h^x)$, *where* $\aleph_h = (\,v, \nabla_x \psi_h\,)$ *and where* a *and* J *are the bilinear operators defined by (3.15) and (3.17).*

If there does exist a discrete DFUG-NIPG solution pair (F_h, ψ_h) to the Vlasov-Poisson system, then it follows by setting $\theta = 0$ that F_h is a discrete DFUG solution to the Vlasov-Poisson system defined by the flow $\aleph_h$, where $\aleph_h = (v, \nabla_x \psi_h)$, i.e., the equations

$$(F_h(\,t=0\,),\, w\,)_{K_j} \ = \ (f_0,\, w\,)_{K_j} \qquad \text{and} \tag{4.154}$$

$$B(\,F_h, w;\, \aleph_h\,) \ = \ - \sum_{f_k \in \Gamma_I} \langle\, f_I\, \aleph_h \cdot \nu_k, w\,\rangle_{f_k}\ , \qquad \forall\ t \in (\,0,T\,], \tag{4.155}$$

are satisfied, $\forall\ w \in D^r(\mathcal{T}_h)$, and it follows by setting $w = 0$ that ψ_h is a discrete NIPG solution to the Poisson system defined by the source term $\rho(F_h)$, i.e., the equation

$$a(\,\psi_h, \theta\,) \ + \ J(\,\psi_h, \theta\,) = (\,\rho(F_h),\, \theta\,)_\Omega\ , \qquad \forall\ t \in (\,0,T\,], \tag{4.156}$$

is satisfied $\forall\ \theta \in D^{\bar{r}}(\mathcal{T}_h^x)$. We also remark that it is easy to show that if piecewise constants are used to approximate f, then the approximate solution will remain nonnegative for all

times, provided the given initial and boundary data for f are nonnegative. This property can be shown by making use of the well-known entropy multiplier test function $w = \beta'(F_h) = \frac{\text{sgn}F_h - 1}{2}$ in the DFUG formulation, where $\beta(F_h) = F_h \frac{\text{sgn}F_h - 1}{2}$ approximates the function $F_h^- :=$*negative part of* F_h. It can then be shown that leads to

$$\frac{d}{dt} \sum_{j=1}^{N_h} \int_{K_j} \beta(F_h)\, dv\, dx \;\leq\; 0\,.$$

If $f_{_0} \geq 0$, then the above condition imlies that $F_h(t) \geq 0,\ \forall\, t \geq 0$.

In this context, we see that F_h is a discrete DFUG solution to the Vlasov system defined by the trio $(\aleph_h, f_{_0}, f_{_I})$ and ψ_h is a discrete NIPG solution to the Poisson system defined by the data pair $(\rho(F_h), r_{_D})$. Hence, we can now use apply all of the previous results for the DFUG and NIPG methods to these perturbed Vlasov and Poisson systems, which is exactly how the upcoming error estimate for the Vlasov-Poisson system is proved.

4.6.2 *A priori* error estimate

We are now ready to present the *a priori* error estimate for the Vlasov-Poisson system. This result will make use of the unique solutions to the perturbed Poisson systems, where such solutions are known to exist by Theorem 10. For the discrete solution pair (F_h, ψ_h), we will denote by $\tilde{\psi}_h$ the unique solution to the Poisson system, having only a Dirichlet boundary condition, defined by the source term $\rho(F_h)$ and the Dirichlet boundary function $r_{_D}$. Theorem 10 guarantees that each solution $\tilde{\psi}_h$ is in $H^1(\Omega)$. However, in the following proof, we will need to assume the stronger regularity condition that $\tilde{\psi}_h \in H^2(\Omega),\ \forall\ h > 0$.

We now state the main theorem of this dissertation.

Theorem 15. *Let $T > 0$ and let $(\, f_{_0}, f_{_I}, r_{_D}\,)$ be a compatible trio that defines the Vlasov-Poisson system (4.20)-(4.25). Assume that this system has a unique classical solution pair (f, ψ) up to time T, according to Definition 1. Moreover, assume that $f \in C^1(\,[0,T],\, H^s(\mathcal{T}_h)\,)$, $s > 3$, and $\psi \in C^1(\,[0,T],\, H^{\bar{s}}(\mathcal{T}_h^x)\,)$, $\bar{s} \geq 2$, are also satisfied. Assume that there exists a unique discrete DFUG-NIPG solution pair $(F_h, \psi_h) \in C^1(\,[0,T],\, D^r(\mathcal{T}_h)\,) \times C^1(\,[0,T],\, D^{\bar{r}}(\mathcal{T}_h^x)\,)$ to this Vlasov-Poisson system. If the true solutions $\{\, \tilde{\psi}_h \,\}_{h>0}$ to the perturbed Poisson systems defined by the source terms $\{\, \rho(F_h) \,\}_{h>0}$ and the Dirichlet boundary function $r_{_D}$ satisfy*

$\tilde{\psi}_h(t) \in C^1(\,[0,T],\, H^{\bar{s}}(\mathcal{T}_h^x)\,),\ \forall\ h>0,$ *and*

$$\sup_{h>0}\ \sup_{0\le t\le T}\ \|\,\tilde{\psi}_h\,\|_{\bar{s},\Omega^x}\ <\ +\infty\,, \tag{4.157}$$

then $\exists$ *a mesh independent constant* C *that scales at most exponentially with the final time* T *such that the following a priori error estimate holds:*

$$\begin{aligned}
&\|\,f(T)-F_h(T)\,\|_{0,\Omega}^2\ +\ \int_0^T \|\!|\,|\,\overline{\aleph_h}\cdot\nu_{_k}\,|^{1/2}\,[\,f-F_h\,]\,|\!\|_{0,\mathring{\mathcal{F}}_h}^2\ +\ \int_0^T \|\!|\,|\,\aleph_h\cdot\nu_{_k}\,|^{1/2}\,(\,f-F_h\,)\,|\!\|_{0,\Gamma_{_O}}^2\\
&+\ \int_0^T \|\!|\,|\,\aleph_h\cdot\nu_{_k}\,|^{1/2}\,(\,f-F_h\,)\,|\!\|_{0,\Gamma_{_I}}^2\\
&\le\ C\left(h^{2\mu-1}\ +\ \frac{h^{2\mu+2s+\bar{\mu}-9}}{r^{2\bar{s}-2}}\ +\ \frac{h^{2\mu+2s+2\bar{\mu}-10}}{r^{2\bar{s}-3}}\ +\ \frac{h^{2\bar{\mu}-4}}{r^{2\bar{s}-3}}\right)\exp(\,C\,h^{4\mu+4s-16}\,)\,.
\end{aligned} \tag{4.158}$$

Moreover, the estimate (4.168) holds upon replacing T *in the lefthandside by any time* $t\in(0,T)$.

Proof. Set the test functions $w=0$ and $\theta=\psi_h$. Then we get that

$$a(\,\psi_h,\psi_h\,)\ +\ J(\,\psi_h,\psi_h\,)\ =\ L(F_h)\,, \tag{4.159}$$

where L is the linear functional defined by (3.24) for the NIPG method. By Theorem 13, we get that

$$\|\,\psi-\psi_h\,\|_{NIPG}^2\ \le\ C\,\|\,\rho(f)-\rho(F_h)\,\|_{0,\Omega^x}^2\ +\ C\,\frac{h^{2\bar{\mu}-2}}{r^{2s-2}}\,\|\,\tilde{\psi}_h\,\|_{s,\Omega^x}^2\,. \tag{4.160}$$

We simplify $\|\rho(f)-\rho(F_h)\|_{0,\Omega^x}^2$ by writing that

$$\begin{aligned}
\|\,\rho(f)-\rho(F_h)\,\|_{0,\Omega^x}^2\ &=\ \int_{\Omega^x}\left(\int_{\Omega^v}(f-F_h)\,dv\right)^2 dx\\
&\le\ \int_{\Omega^x}\left(|\,\Omega^v\,|^{1/2}\left(\int_{\Omega^v}(f-F_h)^2\,dv\right)^{1/2}\right)^2 dx\ =\ |\,\Omega^v\,|\int_{\Omega^x}\int_{\Omega^v}(f-F_h)^2\,dvdx\\
&\le\ C\,\|\,f-F_h\,\|_{0,\Omega}^2\,.
\end{aligned} \tag{4.161}$$

Thus, we get the inequality

$$\| \psi - \psi_h \|^2_{NIPG} \;\leq\; C \| f - F_h \|^2_{0,\Omega} \;+\; C \frac{h^{2\bar{\mu}-2}}{r^{2s-2}} \| \tilde{\psi}_h \|^2_{s,\Omega^x} . \tag{4.162}$$

It follows from this estimate that

$$\| \nabla_x(\psi - \psi_h) \|^2_{0,\Omega} \;\leq\; C \left(\| f - F_h \|^2_{0,\Omega} \;+\; \frac{h^{2\bar{\mu}-2}}{r^{2s-2}} \right) . \tag{4.163}$$

Since ψ_h is a discrete NIPG solution to $\tilde{\psi}_h$, it follows that

$$\begin{aligned} \| [\, \nabla_x \psi_h \,] \|^2_{0,\mathring{\mathcal{G}}_h^x} \;&\leq\; C \frac{h^{2\bar{\mu}-3}}{\bar{r}^{2\bar{s}-3}} \qquad \text{and} \\ \| \nabla_x(\psi - \psi_h) \|^2_{0,\partial\Omega^x} \;&\leq\; C \frac{h^{2\bar{\mu}-3}}{\bar{r}^{2\bar{s}-3}} . \end{aligned} \tag{4.164}$$

Inserting the above three estimates into the estimate in (4.97) given by Theorem 14 , we have that

$$\begin{aligned} &\| f(T) - F_h(T) \|^2_{0,\Omega} \;+\; \int_0^T \| \, | \overline{\aleph_h} \cdot \nu_k \, |^{1/2} [\, f - F_h \,] \|^2_{0,\mathring{\mathcal{F}}_h} \;+\; \int_0^T \| \, | \aleph_h \cdot \nu_k \, |^{1/2} (\, f - F_h \,) \|^2_{0,\Gamma_O} \\ &\;+\; \int_0^T \| \, | \aleph_h \cdot \nu_k \, |^{1/2} (\, f - F_h \,) \|^2_{0,\Gamma_I} \\ &\leq\; C \left(h^{2\mu-1} \;+\; h^{2\mu+2s-8} \int_0^T \left(\| f - F_h \|_{0,\Omega} \;+\; \frac{h^{\bar{\mu}-1}}{r^{s-1}} \right) \right) \\ &\;+\; C \int_0^T \left(h^{2\mu+2s-8} \left(\| f - F_h \|^2_{0,\Omega} \;+\; \frac{h^{2\bar{\mu}-2}}{r^{2s-2}} \right) \;+\; h^{2\mu+2s-7} \frac{h^{2\bar{\mu}-3}}{\bar{r}^{2\bar{s}-3}} \;+\; h^{-1} \frac{h^{2\bar{\mu}-3}}{\bar{r}^{2\bar{s}-3}} \right) . \\ &\leq\; C \left(h^{2\mu-1} \;+\; \frac{h^{2\mu+2s+\bar{\mu}-9}}{r^{2\bar{s}-2}} \;+\; \frac{h^{2\mu+2s+2\bar{\mu}-10}}{r^{2\bar{s}-3}} \;+\; \frac{h^{2\bar{\mu}-4}}{r^{2\bar{s}-3}} \right) \\ &\;+\; C\, h^{2\mu+2s-8} \int_0^T \left(\| f - F_h \|_{0,\Omega} \;+\; \| f - F_h \|^2_{0,\Omega} \right) . \end{aligned} \tag{4.165}$$

We are now ready to complete the proof. Either it is true that

$$\int_0^T \| f - F_h \|_{0,\Omega} \;< \int_0^T \| f - F_h \|^2_{0,\Omega} \tag{4.166}$$

or it is true that

$$\int_0^T \| f - F_h \|_{0,\Omega} \;\geq\; \int_0^T \| f - F_h \|_{0,\Omega}^2 \,. \tag{4.167}$$

If (4.166) is true, then we can apply Gronwall's inequality to (4.165), which results in the desired estimate

$$\begin{aligned}
&\| f(T) - F_h(T) \|_{0,\Omega}^2 \;+\; \int_0^T \| \, | \, \overline{\aleph_h} \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,\mathring{\mathcal{F}}_h}^2 \;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, (\, f - F_h \,) \, \|_{0,\Gamma_O}^2 \\
&\;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, (\, f - F_h \,) \, \|_{0,\Gamma_I}^2 \\
&\leq\; C \left(h^{2\mu-1} \;+\; \frac{h^{2\mu+2s+\bar{\mu}-9}}{r^{2\bar{s}-2}} \;+\; \frac{h^{2\mu+2s+2\bar{\mu}-10}}{r^{2\bar{s}-3}} \;+\; \frac{h^{2\bar{\mu}-4}}{r^{2\bar{s}-3}} \right) \exp(\, C \, h^{4\mu+4s-16} \,) \,.
\end{aligned} \tag{4.168}$$

If, however, (4.167) instead holds, we can apply Lemma 9 to (4.165), which results in the estimate

$$\begin{aligned}
&\| f(T) - F_h(T) \|_{0,\Omega}^2 \;+\; \int_0^T \| \, | \, \overline{\aleph_h} \cdot \nu_k \, |^{1/2} \, [\, f - F_h \,] \, \|_{0,\mathring{\mathcal{F}}_h}^2 \;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, (\, f - F_h \,) \, \|_{0,\Gamma_O}^2 \\
&\;+\; \int_0^T \| \, | \, \aleph_h \cdot \nu_k \, |^{1/2} \, (\, f - F_h \,) \, \|_{0,\Gamma_I}^2 \\
&\leq\; C \left(h^{2\mu-1} \;+\; \frac{h^{2\mu+2s+\bar{\mu}-9}}{r^{2\bar{s}-2}} \;+\; \frac{h^{2\mu+2s+2\bar{\mu}-10}}{r^{2\bar{s}-3}} \;+\; \frac{h^{2\bar{\mu}-4}}{r^{2\bar{s}-3}} \;+\; h^{4\mu+4s-16} \right) \\
&\leq\; C \left(h^{2\mu-1} \;+\; \frac{h^{2\mu+2s+\bar{\mu}-9}}{r^{2\bar{s}-2}} \;+\; \frac{h^{2\mu+2s+2\bar{\mu}-10}}{r^{2\bar{s}-3}} \;+\; \frac{h^{2\bar{\mu}-4}}{r^{2\bar{s}-3}} \;+\; h^{4\mu+4s-16} \right) \exp(C \, h^{4\mu+4s-16}) \,,
\end{aligned} \tag{4.169}$$

where the last follows since $\exp(\, C \, h^{4\mu+4s-16} \,) \;\geq\; 1$, which follows from the fact that $4\mu + 4s - 16 > 0$.

It is important to note that the above proof shows that the constant C either scales linearly or exponentially with the final time T. If we use Gronwall's inequality, which is the case if (4.166) is true, then C scales exponentially with T. If however we use Lemma 9, which is the case when (4.167) is true, then C scales linearly with T. $\square$

4.6.3 Mass/Energy Balance Laws

The mass/energy balance laws of the Vlasov-Poisson system are well-established. The two most common balance laws for this system are those of mass and energy. The particular form these laws take depends on the exact nature of the boundary conditions supplied for the system. In the presentation given herein, we will continue to assume that an inflow boundary condition is given for the Vlasov equation, where the inflow function is identically equal to zero on the boundary of the velocity domain, and that a Dirichlet boundary condition is given for the Poisson equation. Also, we will that there exists a DFUG-NIPG solution to the Vlasov-Poisson system for these given boundary conditions.

If f is a given smooth solution of the Vlasov-Poisson system, then for an arbitrary smooth test function w it follows from integration-by-parts that

$$\int\limits_{\Omega} f_t \, w \, dvdx \; - \int\limits_{\Omega} f \, \alpha \cdot \nabla w \, dxdv \; + \sum_{f_k \in \Gamma_O} \langle \, f w, \, \alpha \cdot \nu_k \, \rangle_{f_k} \; + \sum_{f_k \in \Gamma_I} \langle \, f_I w, \, \alpha \cdot \nu_k \, \rangle_{f_k} \; = \; 0 \, . \tag{4.170}$$

If we set $w = 1$ everywhere in Ω, we get the following conservation of mass law:

$$\frac{d}{dt} \left(\int\limits_{\Omega} f \, dvdx \right) \; + \sum_{f_k \in \Gamma_O} \langle \, f w, \, \alpha \cdot \nu_k \, \rangle_{f_k} \; + \sum_{f_k \in \Gamma_I} \langle \, f_I w, \, \alpha \cdot \nu_k \, \rangle_{f_k} \; = \; 0 \, . \tag{4.171}$$

This equation states that the total mass, or total number of electrons, of the particle system being considered changes only according to the net flux of electrons that entering and exiting the system via the spatial boundary $\partial\Omega^x$.

To derive a discrete analogue to the mass balance law for the DFUG-NIPG solution, we choose the test function $w_h = 1$ everywhere in Ω and we choose the test function $\psi_h = 0$ everywhere in Ω^x. Then from the DFUG-NIPG weak form one easily obtains that

$$\frac{d}{dt} \left(\int\limits_{\Omega} F_h \, dvdx \right) \; + \sum_{f_k \in \Gamma_O} \langle \, F_h \, , \, \aleph_h \cdot \nu_k \, \rangle_{f_k} \; + \sum_{f_k \in \Gamma_I} \langle \, f_I \, , \, \aleph_h \cdot \nu_k \, \rangle_{f_k} \; = \; 0 \, . \tag{4.172}$$

From this identity, we see the discrete mass balance law is nearly identical to the true mass balance law. The only difference is that the true mass balance law can be simplified a bit further by using the fact that the true solution f is equal to zero on that portion of the velocity domain boundary that intersects with Γ_O, whereas the discrete solution is only

known to be close to zero, *i.e.,* within some prescribed error bound, along this boundary

The derivation of the energy balance law for the Vlasov-Poisson system is a bit more involved than was the derivation for the mass balance law. To begin this calculation, we first derive identities for the true kinetic and potential energies, respecively.

To derive an identity for the kinetic energy, we first note that

$$\frac{d}{dt}\left(\int\limits_{\Omega^v} f\,dv\right) \;+\; \nabla_x \cdot \left(\int\limits_{\Omega^v} v\,f\,dv\right) \;+\; \int\limits_{\Omega^v} \nabla_x\psi \cdot \nabla_v f\,dv\,. \tag{4.173}$$

This implies that

$$\rho_t \;+\; \nabla_x \cdot j \;=\; 0\,, \tag{4.174}$$

which follows since $\int_{\Omega^v} \nabla_x\psi \cdot \nabla_v f dv = 0$.

We now multiply the Vlasov equation by the test function $|v|^2$ and integrate the resulting equation over the domain Ω. This leads to

$$\frac{d}{dt}\left(\int\limits_{\Omega} |\,v\,|^2\,f\,dvdx\right) \;+\; \int\limits_{\Omega^v}\int\limits_{\Omega^x} |\,v\,|^2\,\nabla_x\cdot(vf)\,dvdx \;+\; \int\limits_{\Omega^x}\int\limits_{\Omega^v} |\,v\,|^2\,\nabla_v\cdot(\nabla_x\psi\,f)\,dvdx \;=\; 0\,. \tag{4.175}$$

Denoting the current by j and integrating-by-parts, the above equation is rewritten as

$$\frac{d}{dt}\left(\int\limits_{\Omega} |\,v\,|^2\,f\,dvdx\right) \;+\; \int\limits_{\partial\Omega^x}\left(\int\limits_{\Omega^v} |\,v\,|^2\,f\,(v\cdot\nu^x)\,dv\right) dS^x \;-\; 2\int\limits_{\Omega^x} \nabla_x\psi\cdot j\,dx \;=\; 0\,. \tag{4.176}$$

Upon further manipulation, the above identity becomes

$$\begin{aligned}\frac{d}{dt}\left(\int\limits_{\Omega} |\,v\,|^2\,f\,dvdx\right) \;+\; 2\int\limits_{\Omega^x} \psi\cdot\nabla_x j\,dx \;&=\; -\int\limits_{\partial\Omega^x}\left(\int\limits_{\Omega^v} |\,v\,|^2\,f\,(v\cdot\nu^x)\,dv\right) dS^x \\ +\; 2\int\limits_{\partial\Omega^x} r_{_D}\,j\cdot\nu^x\,dS^x\,.\end{aligned} \tag{4.177}$$

We now turn to finding an identity for the potential energy. We do this by using the fact

that

$$\begin{aligned}\frac{1}{2}\,\frac{d}{dt}\left(\int\limits_{\Omega^x}\nabla_x\psi\cdot\nabla_x\psi\,dx\right) &= \int\limits_{\Omega^x}\nabla_x\psi\cdot\nabla_x\psi_t\,dx\\ &= \int\limits_{\partial\Omega^x} r_{_D}\,\nabla_x\psi_t\cdot\nu^x\,dS^x \;-\; \int\limits_{\Omega^x}\psi\,\Delta_x\psi_t\,dx\\ &= \int\limits_{\partial\Omega^x} r_{_D}\,\nabla_x\psi_t\cdot\nu^x\,dS^x \;+\; \int\limits_{\Omega^x}\psi\,\rho_t\,dx\\ &= \int\limits_{\partial\Omega^x} r_{_D}\,\nabla_x\psi_t\cdot\nu^x\,dS^x \;-\; \int\limits_{\Omega^x}\psi\,\nabla_x\cdot j\,dx\,,\end{aligned}$$

which follows from the fact that $\rho_t = -\nabla_x\cdot j$. We rearrange terms to write this identity as

$$2\int\limits_{\Omega^x}\psi\,\nabla_x\,j\,dx \;=\; -\frac{d}{dt}\left(\int\limits_{\Omega^x}|\,\nabla_x\psi\,|^2\,dx\right) \;+\; \int\limits_{\partial\Omega^x} r_{_D}\,\nabla_x\psi_t\cdot\nu^x\,dS^x\,. \tag{4.178}$$

An expression for the rate of change of the total energy is now obtained by substituting (4.178) into (4.177), which results in

$$\begin{aligned}&\frac{d}{dt}\int\limits_{\Omega^x}\left(\int\limits_{\Omega^v}|\,v\,|^2\,f\,dv \;-\; |\,\nabla_x\psi\,|^2\right)dx\\ &= -\int\limits_{\partial\Omega^x}\left(\left(\int\limits_{\Omega^v}|\,v\,|^2\,f\,(v\cdot\nu^x)\,dv\right) \;+\; r_{_D}\left(2\,j \;-\; \nabla_x\psi_t\right)\cdot\nu^x\right)dS^x\,.\end{aligned}\tag{4.179}$$

This equality is the energy balance law for the Vlasov-Poisson system under consideration.

At the discrete level, the potential energy for the NIPG approximation to the potential satisfies

$$\begin{aligned}&\frac{1}{2}\,\frac{d}{dt}\left(\sum_{j=1}^{N_h}\int\limits_{\Omega^x}\nabla_x\psi_h\cdot\nabla_x\psi_h\,dx\right) \;=\; \sum_{j=1}^{N_h}\int\limits_{\Omega^x}\frac{\partial}{\partial t}\,\nabla_x\psi_h\cdot\nabla_x\psi_h\,dx \;=\; \int\limits_{\Omega^x}\frac{\partial}{\partial t}\,\rho_h\,\psi_h\,dx\\ &+\sum_{f^x_{k_x}\in\partial\Omega^x}\nabla_x\psi_h\cdot\nu^x_{k_x}\,\frac{\partial}{\partial t}\,r_{_D}\,dS^x \;-\; \frac{1}{2}\sum_{k=1}^{P_h}\frac{\bar r\sigma_k}{|f^x_{k_x}|}\,\|\,[\psi_h]\,\|^2_{0,f^x_{k_x}} \;+\; \frac{1}{2}\sum_{f^x_{k_x}\in\partial\Omega^x}\frac{\bar r\sigma_k}{|f^x_{k_x}|}\,\|\,r_{_D}-\psi_h\,\|^2_{0,f^x_{k_x}}\end{aligned}$$

Set $w = |\, v \,|^2$. Then we get that

$$\frac{d}{dt} \int\limits_{\Omega} \frac{1}{2} |\, v \,|^2 F_h \, dv \, dx \; - \; \sum_{j=1}^{N_h} \int\limits_{K_j} (\, v \cdot \nabla_x \, \psi_h \,) \, F_h \; + \; \sum_{f_k \in \Gamma_O} \Big\langle \, \frac{1}{2} |\, v \,|^2 F_h \, , \, \aleph_h \cdot \nu_k \, \Big\rangle_{f_k}$$

$$+ \sum_{f_k \in \Gamma_I} \Big\langle \, \frac{1}{2} |\, v \,|^2 f_I \, , \, \aleph_h \cdot \nu_k \, \Big\rangle_{f_k} \; = \; 0 \, . \tag{4.180}$$

Therefore, upon combining the two expressions for the potential and kinetic energies, we get the following discrete energy balance law for the DFUG-NIPG solution pair (F_h, ψ_h) :

$$\frac{d}{dt} \int\limits_{\Omega^x} \left(\int\limits_{\Omega^v} |\, v \,|^2 F_h \, dv \; + \; \sum_{j=1}^{N_h} |\, \nabla_x \psi_h \,|^2 \right) dx \; = \; 2 \int\limits_{\Omega^x} \frac{\partial}{\partial t} \, \rho_h \, \psi_h \, dx$$

$$+ \; 2 \sum_{j=1}^{N_h} \int\limits_{K_j} (\, v \cdot \nabla_x \, \psi_h \,) \, F_h \; - \; 2 \sum_{f^x_{k_x} \in \partial \Omega^x} \nabla_x \psi_h \cdot \nu^x_{k_x} \, \frac{\partial}{\partial t} \, r_{\, D} \, dS^x$$

$$- \sum_{f_k \in \Gamma_O} \Big\langle \, |\, v \,|^2 F_h \, , \, \aleph_h \cdot \nu_k \, \Big\rangle_{f_k} \; - \sum_{f_k \in \Gamma_I} \Big\langle \, |\, v \,|^2 f_I \, , \, \aleph_h \cdot \nu_k \, \Big\rangle_{f_k} \; - \sum_{k=1}^{P_h} \frac{\bar{r} \sigma_k}{|f^x_{k_x}|} \, \| \, [\psi_h] \, \|^2_{0, f^x_{k_x}}$$

$$+ \sum_{f^x_{k_x} \in \partial \Omega^x} \frac{\bar{r} \sigma_k}{|f^x_{k_x}|} \, \| \, r_{\, D} - \psi_h \, \|^2_{0, f^x_{k_x}}$$

$$= \; 2 \int\limits_{\Omega^x} \frac{\partial}{\partial t} \, \rho_h \, \psi_h \, dx \; + \; 2 \sum_{j=1}^{N_h} \int\limits_{K_j} (\, v \cdot \nabla_x \, \psi_h \,) \, F_h \; - \; 2 \sum_{f^x_{k_x} \in \partial \Omega^x} \nabla_x \psi_h \cdot \nu^x_{k_x} \, \frac{\partial}{\partial t} \, r_{\, D} \, dS^x$$

$$- \sum_{f_k \in \Gamma_O} \Big\langle \, |\, v \,|^2 F_h \, , \, \aleph_h \cdot \nu_k \, \Big\rangle_{f_k} \; - \sum_{f_k \in \Gamma_I} \Big\langle \, |\, v \,|^2 f_I \, , \, \aleph_h \cdot \nu_k \, \Big\rangle_{f_k} \; + \; \mathcal{O}(\, \frac{h^{2\bar{\mu}-3}}{r^{2\bar{s}-3}} \,) \, , \tag{4.181}$$

where the last line follows by the previous error estimate theorem. Clearly, this discrete energy balance law is not exact. This results from the fact that some of the terms in (4.181) are error terms. Also, due to the fact that the discrete solution solves a weak formulation, whereas the true solution pair (f, ψ) solves the classical Vlasov-Poisson system, we cannot simplify the discrete law as much as we were able to for the true energy balance equation.

Although the discrete energy balance law is more complicated than the true energy balance law, the error between the total energies for the true system and the discrete system can be estimated in terms of the already established *a priori* error estimate for the DFUG-NIPG

method. To see this, we note that

$$\left| \int_{\Omega^x} \int_{\Omega^v} |\, v \,|^2 \, (\, f \; - \; F_h \,) \, dv \, dx \right| \; \leq \; C \, \| \, f - F_h \, \|_{0,\Omega} \, , \tag{4.182}$$

where here $C = V^2 \, |\, \Omega \,|^{1/2}$, and

$$\begin{aligned} & \sum_{j=1}^{N_h} \int_{K^x_{j_x}} \left(|\, \nabla_x \psi \,|^2 \; - \; |\, \nabla_x \psi_h \,|^2 \right) dx \\ & \qquad = \; \sum_{j=1}^{N_h} \int_{K^x_{j_x}} \left(|\, \nabla_x \psi \,| \; - \; |\, \nabla_x \psi_h \,| \right) \left(|\, \nabla_x \psi \,| \; + \; |\, \nabla_x \psi_h \,| \right) dx \\ & \qquad \leq \; C \, \| \, \psi \; - \; \psi_h \, \|_{0,\Omega^x} \, . \end{aligned} \tag{4.183}$$

Since the explicit estimates are known for the righthandside quantities in (4.182) and (4.183), an explicit error between the total energies for the true and discrete systems can be derived.

4.6.4 Future work on the DFUG-NIPG method of approximation

This theorem is a first attempt at establishing a convergence result for a DG method that approximates the Vlasov-Poisson system. The final estimate is not too pleasing to the eye, especially at first glance. However, the nature of this system is such that any error results established for it are going to be messy. This is in part due to the fact that the distribution f and the potential ψ both have different regularity, and both are approximated by local polynomials of differing orders. This is neccessary, since the convergence results proved in this work demonstrate that the potential needs to be approximated with higher order polynomials than does f, if $h-$optimality is to be maintained.

It remains for future work to try to improve the above theorem. This includes trying to relax the regularity assumptions that were made in the theorem. Also, of paramount importance in our future work is to be able to establish that existence and uniqueness of a discrete DFUG-NIPG solution to the Vlasov-Poisson system. The hope here is that the DFUG-NIPG method is structured in such a way that the procedures that have been developed to prove existence and uniqueness of solutions to the Vlasov-Poisson system, those of using a sequence of converging iterate linear solution pairs, can be adapted for the method. To

this end, the above error estimate is a critical first step, as it establishes many needed results that will most certainly be used in proving an existence and uniqueness result for the discrete solution.

We mention that the DFUG-NIPG method can be put into practice by lagging the discrete flow field in time as follows: given f_0, compute $F_h(t = 0)$ via the DFUG method; using $F_h(t = 0)$, compute $\nabla_x \psi_h(t = 0)$ via the NIPG method; using $\nabla_x \psi_h(t = 0)$, compute $F_h(t = t_1)$, where t_1 is the next time step, using $\nabla_x \psi_h(t = 0)$ and the DFUG method; using $F_h(t = t_1)$, compute $\nabla_x \psi_h(t = t_1)$ via the NIPG method; and etc. Discretizing in time in this way incurs the cost associated with approximating $\nabla_x \psi_h(t = t_{n+1})$ by $\nabla_x \psi_h(t = t_n)$ in the Vlasove equation. However, this approach has the benefit that it linearizes the discrete DFUG-NIPG formulation of the Vlasov-Poisson system.

Chapter 5

Numerical Experiments

In this chapter four numerical examples are presented. In all of the examples, the third-order Runge-Kutta method is used to numerically integrate in time. The first two examples test to see that the correct damping rate is achieved for the linear Landau damping problem for two different equilibrium distributions, where the first example problem was numerically solved for by Cheng and Knorr in 1976 [27] using a finite volume type of method. For convenience, the graph of their damping result is given along with the graph of the damping result computed by using the DFUG-NIPG method. The third benchmark is to compute a numerical solution to the Vlasov-Poisson-Fokker-Planck system to check that the results correspond with currently existing results obtained using other numerical approaches. The last example is to compute a numerical solution to the Vlasov-Poisson system that is subjected to an external force field function for a fixed amount of time to determine if any BGK-like modes are present in the numerical solution.

5.1 Linear Landau damping

Assume that we have a collisionless plasma in two dimensional phase space, where the ions are assumed to be stationary. Then the Vlasov-Poisson system satisfied by the electron

distribution f_1, the electric field E, and the potential ψ is stated as

$$\begin{aligned}
\partial_t f_1 \ + \ v\partial_x f_1 \ - \ E\partial_v f_1 \ &= \ 0 \ , && (5.1)\\
E \ &= \ -\psi_x \ , && (5.2)\\
\psi_{xx} \ &= \ 1 \ - \ \int_{\mathbb{R}} f_1 \, dv \ , && (5.3)\\
f_1(\,0, v, t\,) \ &= \ f_1(\,L, v, t\,), \ \forall \ v \in \mathbb{R}\,, && (5.4)\\
\psi(\,0,\,t\,) \ &= \ \psi(\,L,\,t\,) \ , \qquad \text{and} && (5.5)\\
\psi_x(\,0,\,t\,) \ &= \ \psi_x(\,L,\,t\,) \ , && (5.6)
\end{aligned}$$

for $(x, v, t) \in (0, L) \times (-\infty, \infty) \times (0, \infty)$, where L is some positive integer multiple of 2π, subject to a given initial condition.

Now assume that

$$f_1(x, v, t) \ = \ M(v) \ + \ f(x, v, t), \tag{5.7}$$

where f is a small perturbation compared to $M(v)$ and f is initially L-periodic in x, where $M(v)$ is an equilibrium probability distribution for the electrons. Upon substituting the representation (5.7) into the system (5.1)-(5.3), and then dropping all second- and higher-order terms, we get the following linear system for the perturbation f:

$$\begin{aligned}
\partial_t f \ + \ v\partial_x f \ &= \ E(x,t)\, M'(v) \ , && (5.8)\\
E \ &= \ -\psi_x \ , && (5.9)\\
\psi_{xx} \ &= \ -\int_{\mathbb{R}} f \, dv \ , && (5.10)\\
f(\,0,\,v\,) \ &= \ f(\,L,\,v\,), \ \forall \ v \in \mathbb{R}\,, && (5.11)\\
\psi(\,0,\,t\,) \ &= \ \psi(\,L,\,t\,) \ , \qquad \text{and} && (5.12)\\
\psi_x(\,0,\,t\,) \ &= \ \psi_x(\,L,\,t\,) \ , && (5.13)
\end{aligned}$$

for $(x, v, t) \in (0, L) \times (-\infty, \infty) \times (0, \infty)$, subject to a given initial condition on the perturbation.

In 1946, Landau analyzed the perturbed system (5.8)-(5.13) by first applying the Fourier transform with respect to the physical space variable and then applying the Laplace transform with respect to time. In the case when the equilibrium was chosen to be the Maxwellian and the initial condition for the perturbation is small, Landau showed the Fourier coefficients of the electric field E damped exponentially in time. Specifically, he showed that the

parameters of the $n-th$ coefficient

$$\mathcal{F}(k,w) \;\sim\; e^{i\,(\,kx\,-\,wt\,)}\,, \tag{5.14}$$

where $k = 2\pi n/L$, satisfy the dispersion relation

$$\epsilon(k,w) \;=\; 0\,, \tag{5.15}$$

where $\epsilon(k,w)$ is defined by

$$\epsilon(k,w) \;=\; 1 \;-\; \frac{w^2}{k^2}\int\limits_{\mathbb{R}} \frac{M'(v)}{v-\frac{w}{k}}\,dv\;, \qquad (\,\frac{w}{k}\in UHP\,)\,.$$

The significance of the dispersion relation (5.15) is that it solutions characterized the asymptotic-time behavior of the coefficient $\mathcal{F}(k,w)$. In particular, if $w = w_{_R} + i\gamma$, where $w_{_R}$ and γ are real-valued, then γ is the rate of time decay, since $\mathcal{F}(k,w) \sim e^{-iwt}$ in time.

We now present two numerical examples for the linear Landau damping problem, one for the case when the equilibrium is the Maxwellian distribution, which was treated by Landau, and one for the case when the equilibrium is the Lorenztian distribution. For the example involving the Lorenztian distribution, the function $w = w(k)$ will be derived by solving the dispersion relation.

5.1.1 Example 1: Maxwellian equilibrium

Define the equilibrium to be the Maxwellian distribution

$$M(v) \;=\; \frac{1}{2\pi}\,e^{-\frac{v^2}{2}}\,. \tag{5.16}$$

For this equilibrium function, Landau showed that the exponential decay rate $\gamma_{_{theor}}$ of the Fourier mode corresponding to the wave number k, was approximately equal to

$$\gamma_{_{theor}}(k) \;\approx\; -\sqrt{\frac{\pi}{8}}\,\frac{1}{k^3}\exp\big(-\frac{1}{2k^2}-\frac{3}{2}\big)\,. \tag{5.17}$$

As a benchmark test for DFUG-NIPG method, equations (5.8)-(5.13) be approximated by using the DFUG-NIPG method. The numerical decay rate $\gamma_{_{num}}$ resulting from the approximation will be compared against (5.17) as a check for accuracy. This will be performed by choosing appropriate parameter values in (5.8)-(5.13) so that a specific Fourier mode is

excited. Then, by the periodic nature of this system, this mode will remain the dominate mode for all times. Thus, the Fourier coefficient of the electric field corresponding to this mode will determine the behavior of E as it evolves over time.

The exact system that will be numerically approximated was investigated by Cheng and Knorr in [27] and is stated as follows:

$$\begin{aligned} f_t \; + \; v\, f_x &= E(x,t)M'(v)\;, && (5.18)\\ E &= -\psi_x\;, && (5.19)\\ \psi_{xx} &= -\int_{-\infty}^{\infty} f\, dv\,, && (5.20)\\ f(x,v,0) &= \epsilon\cos(kx)\, M(v)\;, && (5.21)\\ f(0,v,t) &= f(L,v,t) && (5.22)\\ \psi(0,t) &= \psi(L,t) = 0\;, && (5.23) \end{aligned}$$

for $(x,v,t) \in (0,L)\times(-\infty,\infty)\times(0,\infty)$, where $\epsilon = 0.01,\ k = 0.5$, and $L = 4\pi$. The given parameter values for k and L correspond to the mode $n = 1$. Using the above approximate formula for γ_{theor} with a slight correction in the formula to improve its accuracy, we get that

$$\gamma_{theor}(k) \;\approx\; -0.153\;. \qquad (5.24)$$

Thus, for this example, we will verify that the numerical approximation to the field damps exponentially with a rate that is close to -0.153.

Mesh details

The system (5.18)-(5.19) was computed using DFUG-NIPG method on a mesh comprised of 70 uniform elements in the $x-$direction and 260 nonuniform elements in the $v-$direction. The velocity domain was set to $[-6,6]$. The elements in the $v-$direction were spaced as follows: 10 unifomly spaced elements from $[-6,-4]$, 240 uniformly spaced elements from $[-4,4]$, and 10 unifomly spaced elements from $[4,6]$. Tensor-product quadratic polynomials of degree two were used to define the approximation spaces for $f(x,v,t)$ and $E(x,t)$. The numerical results are graphed in Figures 5.1 and 5.2.

Discussion of Results

The plot in Figure 5.2 shows that the electric field damps exponentially in time. To compute the numerical damping rate, the slopes of each of the line segments connecting the local maximums in time of the logarithm of the spatial L^∞-norm of the electric field function, *i.e.*, $\log \|\nabla_{\mathrm{x}}\psi_{\mathrm{h}}(\mathrm{t})\|_{\mathrm{L}^\infty(\Omega^{\mathrm{x}})}$, are calculated. The average value of all of these slopes is then computed and γ_{num} is defined to be this average value. In this example, we compute that $\gamma_{num} = -0.153$.

The difficulty in capturing the Landau damping effect numerically is due to the fact that the filamentation becomes more severe as time grows large. The nature of the filamentation for large times can be seen in Figure 5.1. There are two key advantages in using DG to approximate this system. First, the ability to easily employ a nonuniform mesh, which is the case for any DG method, allows one use many elements in those regions of the domain where the function under consideration experiences rapid variations, which is the case here. In those regions where the function experiences only minor variations, coarse elements can be used. The second key advantage of the DG method is that the approximate solution is not forced to continuous across inter-element boundaries. This is important fact when approximating functions having rapid oscillations, since a discrete DG solution is better able to resolve oscillations than a CG method is able to do. For Landau damping problems, the fact that a CG method would smooth out oscillations, perhaps while still preserving mass, would result in the method not be able to capturing the damping effect, since the presence of this effect is due to filamentation.

5.1.2 Example 2: Lorenztian equilibrium

Define the equilibrium to be the Lorentzian distribution

$$M(v) = \left(\frac{\pi}{v_\theta}\right)^{-1} \frac{1}{v^2 + v_\theta^2}, \tag{5.25}$$

where $v_\theta \in \mathbb{R}$. This distribution has much slower decaying tails than the Maxwellian distribution, which is what makes it of interest. The same arguments that Landau used for the Maxwellian equilibrium can be used for the Lorentzian equilibrium, with a few modifications. The great thing about using the Lorentzian equilibrium is that it turns out that the exponential decay rate γ_{theor} of the Fourier mode corresponding to the wave number k can be exactly computed.

Given the definition of M above, it follows that

$$M'(v) \;=\; -2\left(\frac{\pi}{v_\theta}\right)^{-1} \frac{v}{(\,v^2 \;+\; v_\theta^2\,)^2}\;. \tag{5.26}$$

This implies $\epsilon(k,w)$ is equal to

$$\epsilon(k,w) \;=\; 1 \;+\; \frac{2}{k^2}\left(\frac{\pi}{v_\theta}\right)^{-1} \int\limits_{\mathbb{R}} \frac{v}{(\,v^2 \;+\; v_\theta^2\,)^2\,(\,v-\frac{w}{k}\,)}\,dv\;, \qquad (\,\frac{w}{k} \in UHP\,)\,. \tag{5.27}$$

The above integral can be simplified by first writing that

$$\begin{aligned}\int\limits_{\mathbb{R}} \frac{v}{(\,v^2 \;+\; v_\theta^2\,)^2\,(\,v-\frac{w}{k}\,)}\,dv \;&=\; \int\limits_{\mathbb{R}} \frac{dv}{(\,v \;-iv_\theta\,)^2\,(\,v \;+\; iv_\theta\,)^2} \\ &\quad+\; \frac{w}{k}\int\limits_{\mathbb{R}} \frac{dv}{(\,v \;-iv_\theta\,)^2\,(\,v \;+\; iv_\theta\,)^2\,(\,v-\frac{w}{k}\,)} \\ &=:\; N \;+\; Y\left(\frac{w}{k}\right)\,.\end{aligned}$$

Now we recall that Cauchy's formula for higher-order poles states that if g is an analytic function in a domain D containing a piecewise smooth, simple, closed, directed curve C, then for any point $c \in C$ we have that

$$2\pi i\,\frac{d^n}{d\beta^n}\,g(c) \;=\; \int\limits_C \frac{g(\beta)}{(\beta-c)^{n+1}}\,d\beta\,.$$

Using this formula, we easily obtain that

$$N \;=\; 2\pi i\,\frac{d}{dv}\left(\frac{1}{(v \;+\; iv_\theta)^2}\right)_{|_{v=iv_\theta}} \;=\; 2\pi i\left(\frac{-2}{(v \;+\; iv_\theta)^3}\right)_{|_{v=iv_\theta}} \;=\; \frac{\pi}{2\,v_\theta^3}\,. \tag{5.28}$$

Let $u = w/k$. Then to compute $Y\left(\frac{w}{k}\right) = Y(u)$, we make use of the fact that $u = \frac{w}{k} \in UHP$.

From Cauchy's formula, it follows that

$$\begin{aligned} Y(u) &= u \int\limits_{\mathbb{R}} \frac{1}{(\,v\, - iv_\theta\,)^2}\,\frac{1}{(\,v\, +\, iv_\theta\,)^2\,(\,v-u\,)}\,dv \\ &= 2\pi i\,u\,\frac{d}{dv}\,\left(\frac{1}{(v\, +\, iv_\theta)^2\,(v-u)}\right)_{|_{v=iv_\theta}} \\ &= 2\pi i\,u\,\frac{d}{dv}\,\left(\frac{-2}{(v+iv_\theta)^3\,(v-u)}\, -\, \frac{1}{(v+iv_\theta)^2\,(v-u)^2}\right)_{|_{v=iv_\theta}} \\ &= \frac{-\pi u}{2v_\theta^3}\,\frac{(\,u-2iv_\theta\,)}{(\,u-iv_\theta\,)^2}\;=\; -uN\,\frac{(\,u-2iv_\theta\,)}{(\,u-iv_\theta\,)^2}\;. \end{aligned}$$

Combining the identities for N and Y, we get that

$$\begin{aligned} \epsilon(k,u) &= 1\, +\, \frac{2}{k^2}\,\left(\frac{\pi}{v_\theta}\right)^{-1}\,(\,N\, +\, Y(u)\,) \\ &= 1\, +\, \frac{2}{k^2}\,\left(\frac{\pi}{v_\theta}\right)^{-1}\,\left(N\, -\, uN\,\frac{(\,u-2iv_\theta\,)}{(u-iv_\theta)^2}\right) \\ &= 1\, +\, \frac{2N}{k^2}\,\left(\frac{\pi}{v_\theta}\right)^{-1}\,\left(1\, -\, \frac{(\,u^2-2iuv_\theta-v_\theta^2+v_\theta^2\,)}{(u-iv_\theta)^2}\right) \\ &= 1\, +\, \frac{2N}{k^2}\,\left(\frac{\pi}{v_\theta}\right)^{-1}\,\left(1\, -\, 1\, -\, \frac{v_\theta^2}{(u-iv_\theta)^2}\right) \\ &= 1\, -\, \frac{2N}{k^2}\,\left(\frac{\pi}{v_\theta}\right)^{-1}\,\frac{v_\theta^2}{(u-iv_\theta)^2}\;. \end{aligned}$$

Using the above identity for ϵ, we can now solve the dispersion relation (5.15). Upon setting $A = (\pi/v_\theta)^{-1}$, we see that the dispersion relation is equivalent to

$$\begin{aligned} & \frac{2N}{k^2}\,A\,\frac{v_\theta^2}{(u-iv_\theta)^2}\;=\;1 \\ \Leftrightarrow\quad & 2N\,A\,v_\theta^2\;=\;k^2\,(\,u-iv_\theta\,)^2 \\ \Leftrightarrow\quad & w^2\;-\;(2ikv_\theta)\,w\;-\;(2AN\;+\;k)\,v_\theta^2\;=\;0\,. \end{aligned}$$

The two solutions to the above quadratic equation in w are

$$w\;=\;v_\theta\,\sqrt{2AN+k-k^2}\;+\;i\,(kv_\theta) \qquad \text{and} \qquad w\;=\;-v_\theta\,\sqrt{2AN+k-k^2}\;+\;i\,(kv_\theta)\,.$$

Therefore, for a Lorentzian case, we end up with the formula

$$\gamma_{theor}(k, v_\theta) \;=\; k\, v_\theta \,. \tag{5.29}$$

We note that this formula is exact, whereas the formula for the Maxwellian case was an approximation.

The exact problem that will be numerically approximated is as follows:

$$\begin{aligned}
f_t \;+\; v\, f_x &= E(x,t)M'(v)\;, && (5.30)\\
E &= -\psi_x\;, && (5.31)\\
\psi_{xx} &= -\int_{-\infty}^{\infty} f\, dv\,, && (5.32)\\
f(x,v,0) &= \epsilon\cos(kx)\, M(v)\;, && (5.33)\\
f(0,v,t) &= f(L,v,t) && (5.34)\\
\psi(0,t) &= \psi(L,t) = 0\;, && (5.35)
\end{aligned}$$

for $(x, v, t) \in (0, L) \times (-\infty, \infty) \times (0, \infty)$, where $\epsilon = 0.01,\ k = 0.25,\ v_\theta = 1$ and $L = 8\pi$. The wave number $k = 0.25$ corresponds to the mode $n = 4$ in this problem. Using the above exact formula we get that

$$\gamma_{theor}(k) \;= k\, v_\theta \;=\; 0.25\;. \tag{5.36}$$

Thus, in this problem, we will verify that the numerical approximation to the field damps exponentially with a rate near -0.25.

Mesh Details

The system (5.30)-(5.35) was computed using DFUG-NIPG method on a mesh comprised of 120 uniform elements in the $x-$direction and 360 nonuniform elements in the $v-$direction. The velocity domain was set to $[-20, 20]$. The elements in the $v-$direction were spaced as follows: 20 unifomly spaced elements from $[-20, -10]$, 320 uniformly spaced elements from $[-10, 10]$, and 20 unifomly spaced elements from $[10, 20]$. Tensor-product quadratic polynomials of degree two were used to define the approximation spaces for $f(x, v, t)$ and $E(x, t)$. The numerical results are graphed in Figures 5.3.

Discussion of Results

The Landau damping problem for the Lorenztian distribution is much more challenging than the corresponding problem for the Maxwellian distribution. This mainly results from the fact that the computational velocity domain must be much larger for the Lorentzian case, since the tails of the Lorentzian decay much more slowly in v than those for the Maxwellian. Thus, we expect that the Landau damping effect will not be captured in this example as long as it was captured in the first example problem. This is indeed the case, as can be seen from looking at Figure 5.3. The damping effect is captured for only a third of the time that it is captured for the Maxwellian case.

5.2 Example 3: Schematic of channel region of semiconductor device

This example problem was numerically solved using the Weighted Essentially Nonoscillatory (WENO) method in [1], where the exact details of this problem were clearly given. For comparitive purposes in this paper, results were also presented that were generated using a direct-simulation Monte-Carlo (DSMC) method. In this problem, the system being computed is the Vlasov-Poisson-Fokker-Planck system. The DFUG-NIPG method is used to discretize both the Vlasov-Poisson system and the NIPG is used to discretize the Fokker-Planck operator.

The precise problem statement for this example is the following: find (f, E, ψ) such that the system

$$f_t \; + \; v f_x \; - \; \frac{e}{m} E \, f_v \;\; = \;\; \frac{1}{\tau} \, (\, \theta \, f_v \; + \; v \, f \,)_v \; , \tag{5.37}$$

$$E \;\; = \;\; -\psi_x \; , \tag{5.38}$$

$$\psi_{xx} \;\; = \;\; \frac{e}{\epsilon_0} \left(\int_{-\infty}^{\infty} f \, dv \; - \; C(x) \right) \; , \tag{5.39}$$

$$f(x, v, 0) \;\; = \;\; C(x) \, M(v) \; , \tag{5.40}$$

$$f(0, v, t) \;\; = \;\; C(0) \, M(v) \; , \tag{5.41}$$

$$f(L, v, t) \;\; = \;\; C(L) \, M(v) \; , \tag{5.42}$$

$$\psi(0, t) \;\; = \;\; 0 \text{ volts} \; , \tag{5.43}$$

$$\psi(L, t) \;\; = \;\; 2 \text{ volts} \, , \tag{5.44}$$

is satisfied for $(x, v, t) \in (0, L) \times (-\infty, \infty) \times (0, \infty)$, where $M(v)$ is the Maxwellian distribution. The doping function $C(x)$ is the same as that used in [1] and the algorithm for generating this function was provided by the authors of this reference. A plot of the doping function is given in Figure 5.4. All of the above parameters were set to those values given in [1].

Mesh Details

We fix $L = 0.6$. The velocity domain is set to be $[-4, 4]$. The mesh used for this computation consists of 60 elements in the $x-$direction and 100 elements in the $v-$direction. The mesh is nonuniform in the velocity direction. It employs 80 uniform elements to partition $[1.8, 1.8]$ and it uses 10 uniform elements on each of the intervals $[-4, -1.8]$ and $[1.8, 4]$. Tensor-product quadratic polynomials of degree two were used to define the approximation spaces for $f(x, v, t)$ and $E(x, t)$. The numerical results are are graphed in Figures 5.4 - 5.9.

Discussion of Results

The results of this example correspond extremely well with those reported in [1]. The plots of the potential, the electric field, the density, and the current given in Figures 5.6 - 5.9 appear nearly identical to those given in [1].

5.3 Example 4: Laser-plasma interaction (KEEN waves)

In this example, we compute, using the DFUG-NIPG method, the numerical solution to the one dimensional Vlasov-Poisson system that is subjected to the external pondermotive force field function $E_{_D}(x, t)$ for a fixed period of time. This particular example was introduced and computed by Afeyan *et al.* [3].

The pondermotive external force field function is defined to be

$$E_{_D}(x,t) \;=\; \begin{cases} A_{_D}(t)\sin(\,k\,x \;-\; w\,t\,) & , \quad \text{for } \; 0 \le t \le 120\,, \\ 0 & , \quad \text{for } \; t > 120\,. \end{cases}$$

A plot of this function is given in Figure 5.10.

The exact problem statement is to find (f, E, ψ) such that the system

$$
\begin{aligned}
f_t \; + \; v\, f_x \; - \; \frac{e}{m}\,(\,E \; + \; E_{_D}\,)\, f_v &= 0\,, && (5.45)\\
E &= -\psi_x\,, && (5.46)\\
-\psi_{xx} &= 1 \; - \; \int_{-\infty}^{\infty} f\, dv\,, && (5.47)\\
f(x=0) &= f(x=L) \qquad \text{on the inflow boundary}\,, && (5.48)\\
f(x,v,0) &= M(v)\,, && (5.49)\\
f(0,v,t) &= f(L,v,t)\,, && (5.50)\\
\psi(0,t) &= \psi(L,t) = 0\,, && (5.51)
\end{aligned}
$$

is satisfied for $(x, v, t) \in (0, L) \times (-\infty, \infty) \times (0, \infty)$.

Mesh Details

The above system was computed using DFUG-NIPG method on a mesh comprised of 160 uniform elements in the $x-$direction and 320 nonuniform elements in the $v-$direction. The spatial domain was set to $[-4\pi, 4\pi]$ and the velocity domain was set to $[-9, 9]$. The elements in the $v-$direction were spaced as follows: 20 unifomly spaced elements from $[-9, -4]$, 280 uniformly spaced elements from $[-4, 4]$, and 20 unifomly spaced elements from $[4, 9]$. Tensor-product quadratic polynomials of degree two were used to define the approximation spaces for $f(x, v, t)$ and $E(x, t)$. The numerical results are graphed in Figures 5.10 - 5.12.

Discussion of Results

These results show the presence of an electron hole and the filamentation that results from the nature of the Vlasov-Poisson system. The cross-sectional plots given in Figure 5.12 show that approximate solution F_h is in fact discontinuous in those regions experiencing rapid variations. Clearly, the electon hole is dissipating in time. Whether or not this dissipation is due to physical reason or results from the DFUG-NIPG scheme remains an open question.

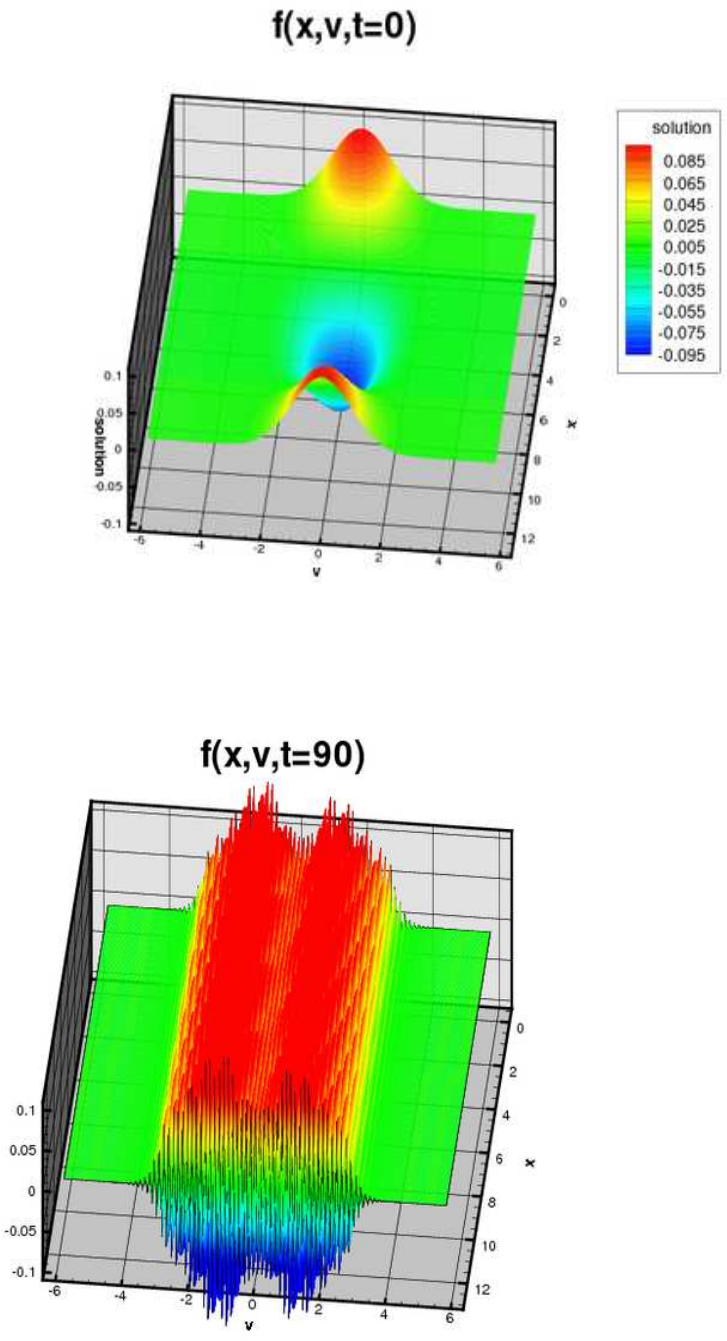

FIGURE 5.1: *(Linear Landau Damping) Top: Plot of initial perturbation. Bottom: Plot of perturbation at t=90.*

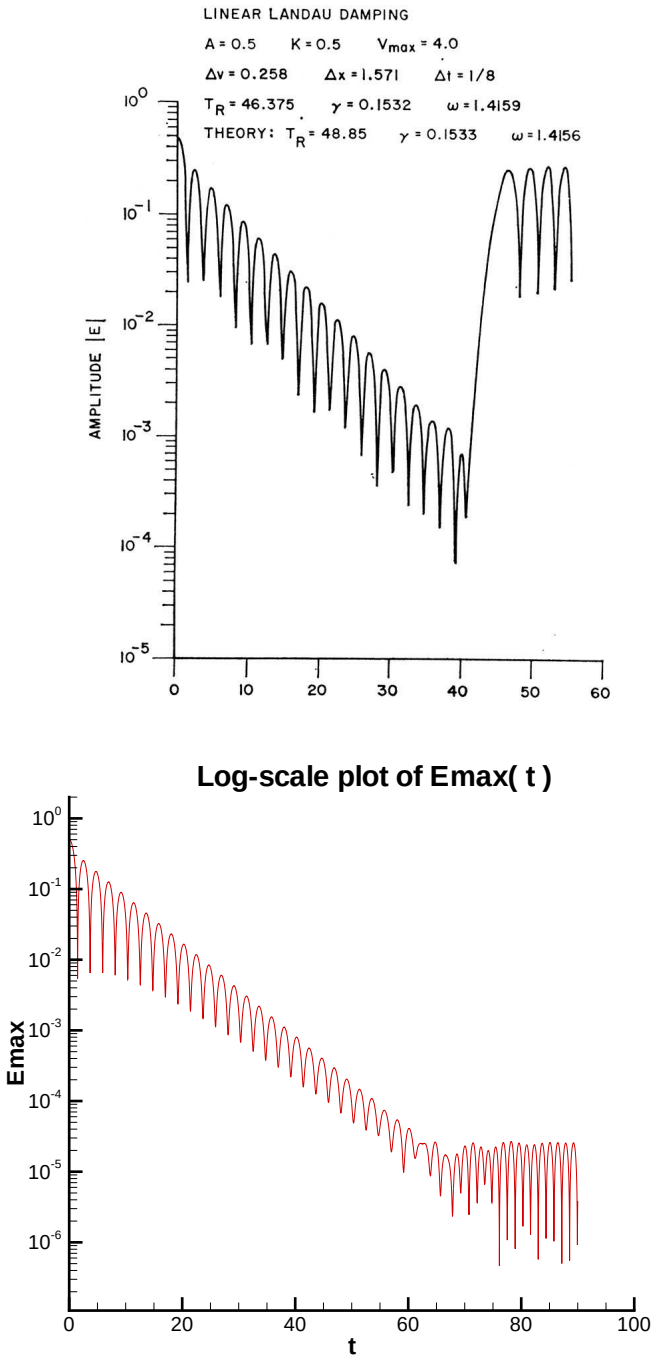

FIGURE 5.2: *Linear Landau damping with* $\triangle x = 0.179$, $\triangle v = 0.033$, $\triangle t = 0.0005$, $\gamma_{num} = -0.153$, *and* $\gamma_{theor} = -0.153$, *where* $\mathrm{Emax}(t) := \|E(\cdot, t)\|_{L^\infty(0,4\pi)}$. *Top: Original damping result of Cheng and Knorr. Bottom: Damping plot result using DFUG-NIPG method.*

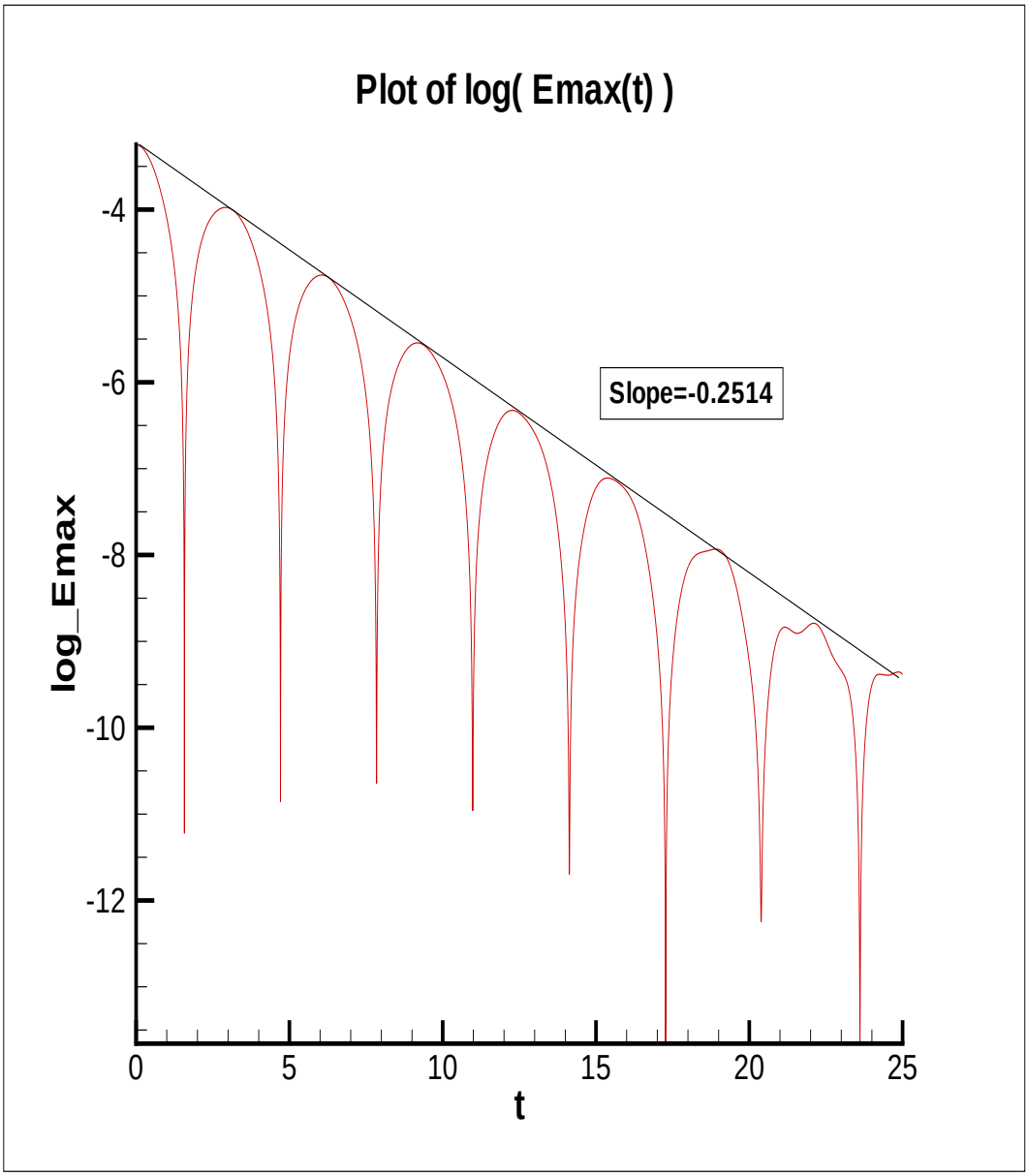

FIGURE 5.3: *Linear Landau damping with* $\triangle x = 0.028$, $\triangle v = 0.0625$, $\triangle t = 0.0005$, $\gamma_{num} = -0.2514$, *and* $\gamma_{theor} = -0.25$, *where* $\text{Emax}(t) := \|E(\cdot,t)\|_{L^\infty(0,8\pi)}$. *Damping plot result for Lorentzian equilibrium using DFUG-NIPG method.*

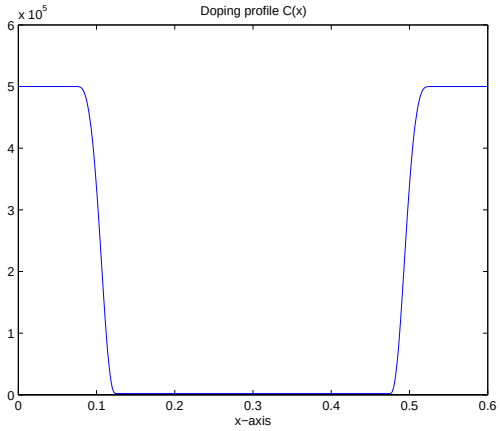

FIGURE 5.4: *Plot of the doping profile function $C(x)$ for Vlasov-Poisson-Fokker-Planck semiconductor problem.*

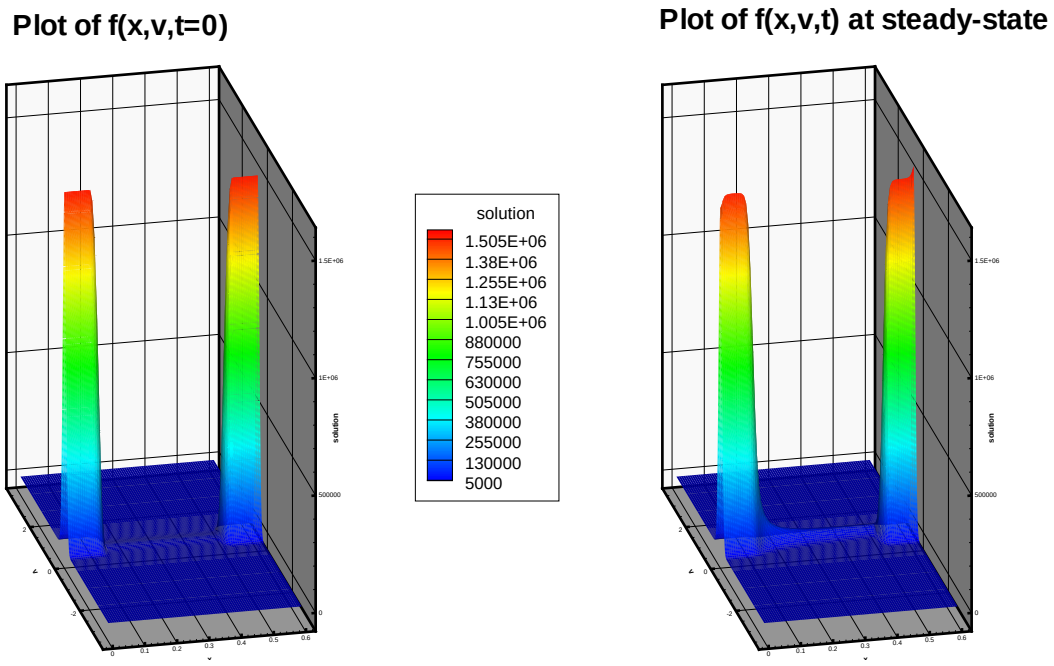

FIGURE 5.5: *Doping profile $C(x)$ and initial and final plots of the distribution $f(x, v, t)$ for Vlasov-Poisson-Fokker-Planck semiconductor problem.*

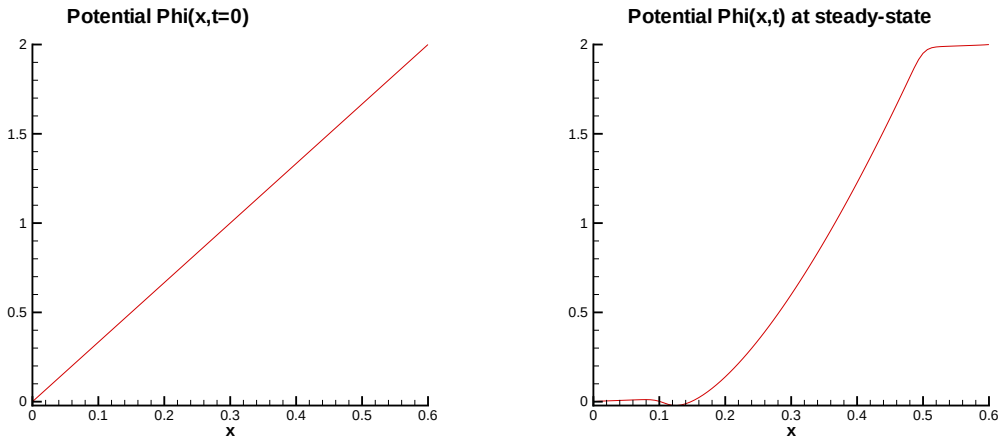

FIGURE 5.6: *Initial and final plots of the potential $\psi(x,t)$ for Vlasov-Poisson-Fokker-Planck semiconductor problem.*

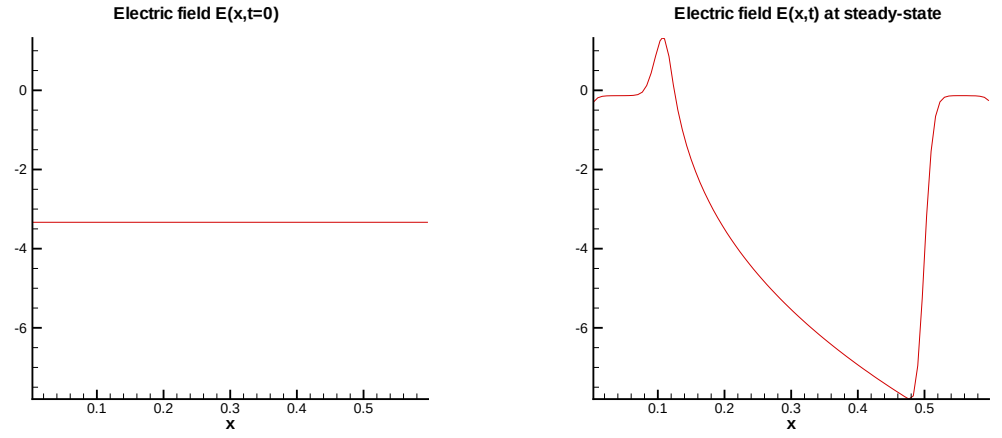

FIGURE 5.7: *Initial and final plots of the electric field $E(x,t)$ for Vlasov-Poisson-Fokker-Planck semiconductor problem.*

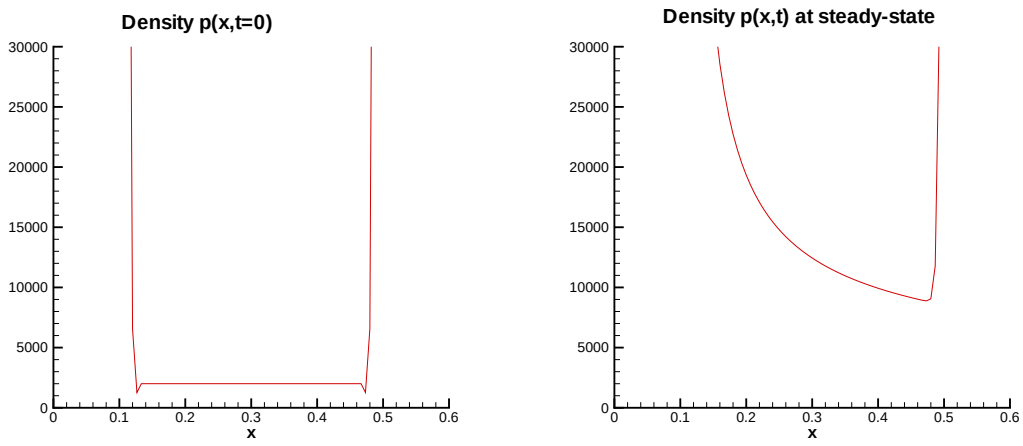

FIGURE 5.8: *Initial and final plots of the density $\rho(x,t)$ for Vlasov-Poisson-Fokker-Planck semiconductor problem.*

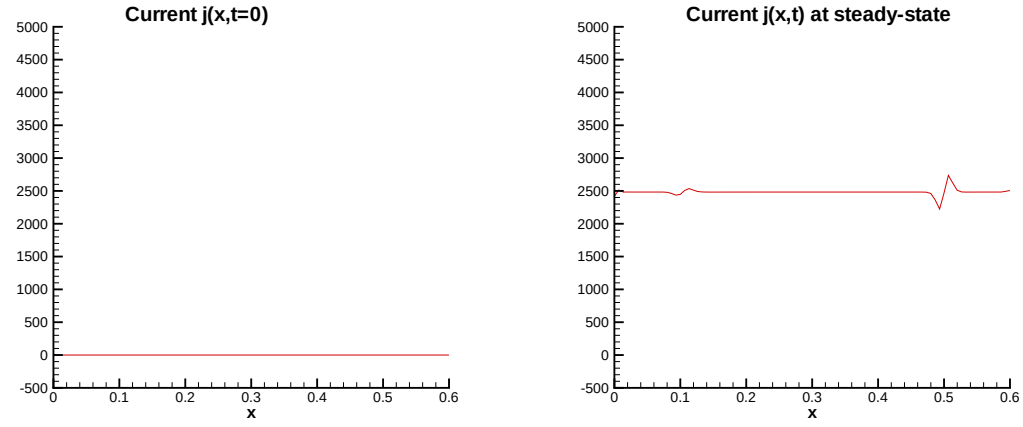

FIGURE 5.9: *Initial and final plots of the current $j(x,t)$ for Vlasov-Poisson-Fokker-Planck semiconductor problem.*

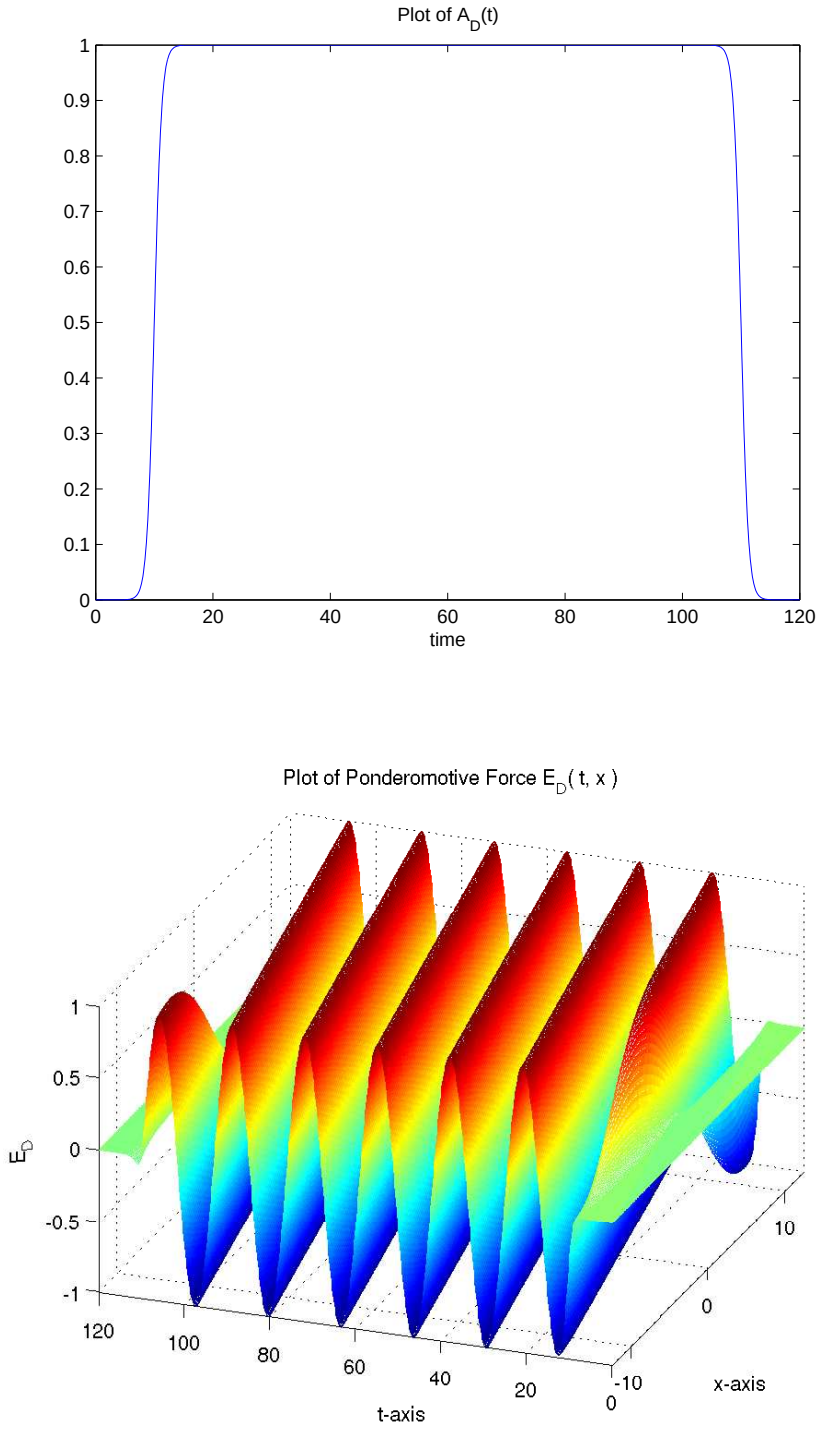

FIGURE 5.10: *Top: Plot of the ramping function $A_D(t)$. Bottom: Plot of the pondermotive forcing function $E_D(x,t) = A_D(t)$.*

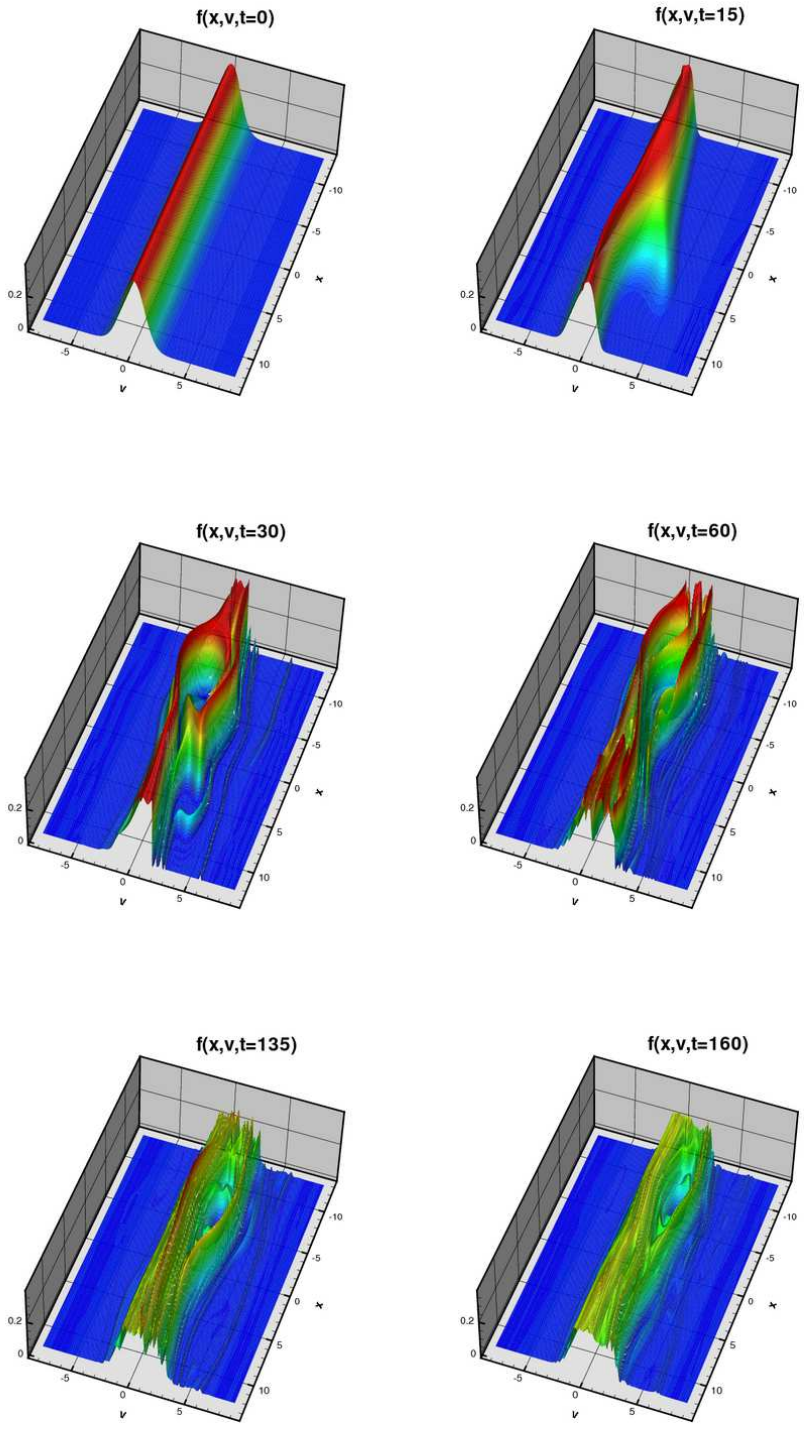

FIGURE 5.11: *Plots of the solution* $f(x, v, t)$ *at times* $t = 0$, 15, 30, 60, 135, *and* 160.

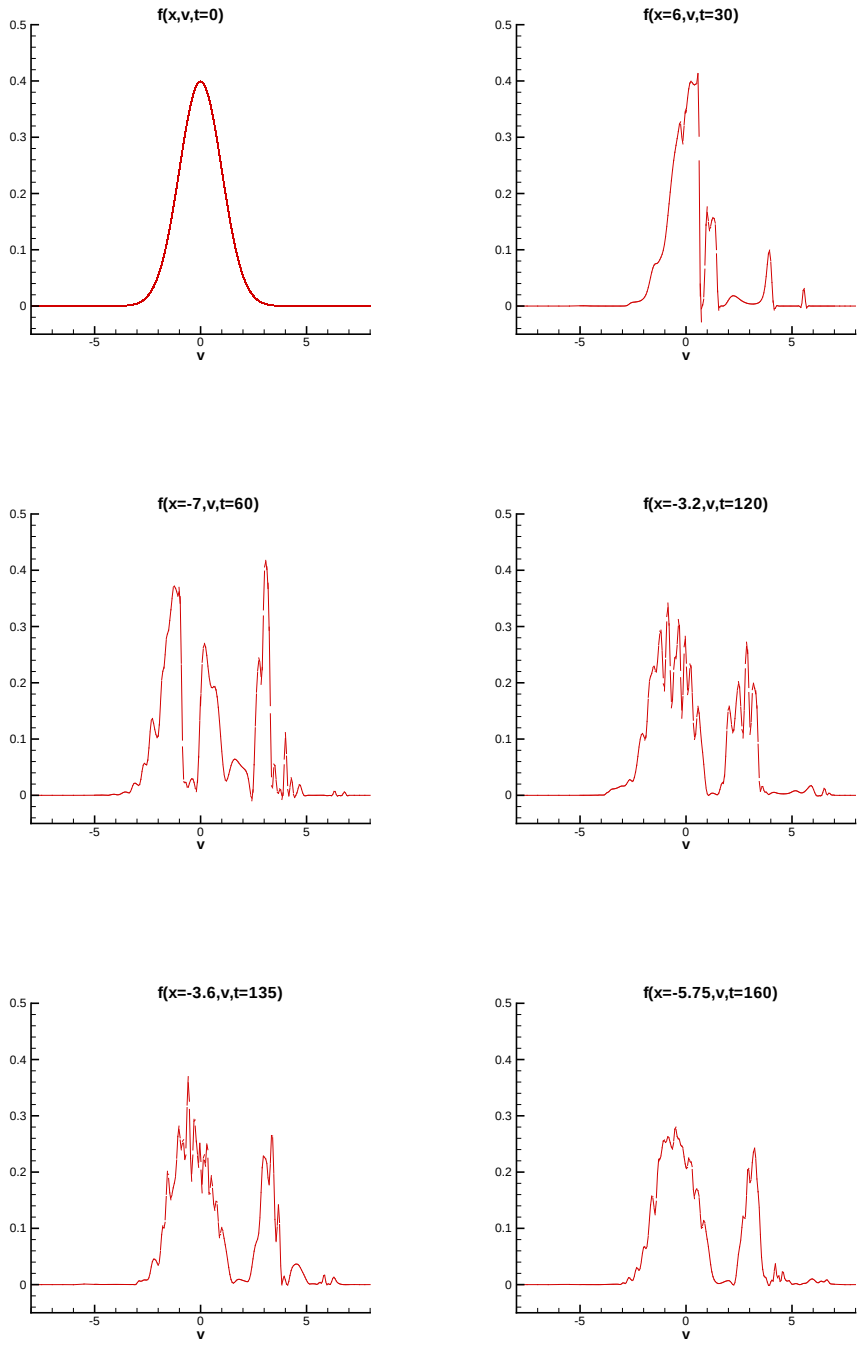

FIGURE 5.12: *Cross-sectional plots, x is fixed, of the solution $f(x, v, t)$ at times $t =$* 0, 15, 30, 60, 135, *and* 160. *The values that x is fixed at are chosen so that the cross-sections slice through the middle of the electron hole.*

Chapter 6

Conclusions and Future Work

The initial theoretical and numerical results presented in this dissertation indicate that the discontinuous Galerkin method is suitable for approximating collisionless plasmas modeled by the Vlasov-Poisson system. Based on the results of this work, the future directions of research are the following:

- extend the NIPG error analysis for the perturbed source term problem to include the use of Neumann boundary conditions,
- relax the regularity requirements in the convergence proof for the DFUG method applied to the perturbed Vlasov system,
- relax the regularity requirments in the convergence proof for the DFUG-NIPG method applied to the Vlasov-Poisson system and relax the Dirichlet boundary condition required in this proof as well,
- prove existence and uniqueness of a discrete solution to the DFUG-NIPG method,
- develop a high-order explicit time integrator to linearize the nonlinear DFUG-NIPG discretization of the Vlasov-Poisson system,
- extend 2D phase space code to more complicated geometries and boundary conditions,
- test code results against theoretical results, especially Landau damping arising from perturbations about more complicated equilibriums,
- extend code to higher dimensional domains.

Bibliography

[1] A.Arnold, J.A. Carillo, I.M. Gamba, and C.W. Shu. Low and high-field scaling limits for the Vlasov- and Wigner-Poisson-Fokker-Planck systems. *Transport Theory and Statistical Physics*, 3:121–153, 2001.

[2] R.A. Adams. *Sobolev Spaces.* Academic Press, New York City, 1975.

[3] B. Afeyan, K. Won, V. Savchenko, T.W. Johnston, A. Ghizzo, and P. Bertrand. A new plasma physics periodic kinetic electrostatic electron nonlinear (KEEN) excitation and its relation to SEAS (simulated elctron acoustic scattering). 31st Conference of Plasma Physics, June 28 - July 2 2004.

[4] V. Aizinger, C. Dawson, B. Cockburn, and P. Castillo. Local discontinuous Galerkin method for contaminant transport. *Advances in water resources*, 24:73–87, 2000.

[5] T. Armstrong, R. Harding, G. Knorr, and D.C. Montgomery. Solution of Vlasov's equation by transform methods. In B.J. Alder, S. Lembach, and M. Rotenberg, editors, *Methods in Computational Physics*, volume 9, pages 30–86. Academic Press, 1970.

[6] A. Arnold, J. A. Carillo, L. Desvillettes, J. Dolbeault, A. Jungell, C. Lederman, P. A. Markowich, G. Toscani, and C. Villani. Entropies and equilbria of many-particle systems: an essay on recent research. *Monatshefte fur Mathematik*, 142:35–43, 2004.

[7] D.N. Arnold. *An Interior Penalty Finite Element Method with Discontinuous Elements.* PhD thesis,The University of Chicago, Chicago, 1979.

[8] D.N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM Journal on Numerical Analysis Numerical Analysis*, 19(4):742–760, 1982.

[9] D.N. Arnold, F. Brezzi, B. Cockburn, and L.D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM Journal on Numerical Analysis Numerical Analysis*, 39(5):1749–1779, 2002.

[10] I. Babuska. The finite element method with Lagrangian multipliers. *Numer. Math.*, 20:179–192, 1973.

[11] I. Babuska. The finite element method with penalty. *Math. Comp.*, 27(122):221–228, 1973.

[12] I. Babuska and M. Suri. The h-p version of the finite element method with quasi-uniform meshes. *RAIRO. Mathematical Modelling and Numerical Analysis*, 21:199–238, 1987.

[13] I. Babuska and M. Suri. The optimal convergence rates of the p version of the finite element method. *SIAM Journal on Numerical Analysis Numerical Analysis*, 24(4):750–776, 1987.

[14] I. Babuska and M. Zlamal. Nonconforming elements in the finite element method with penalty. *SIAM Journal on Numerical Analysis Numerical Analysis*, 10:863–875, 1973.

[15] F. Bassi and S. Rebay. A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations. *Journal of Computational Physics*, 131:267–279, 1982.

[16] C.E. Baumann. *An hp-Adaptive Discontinuous Finite Element Method for Computational Fluid Dynamics*. PhD thesis,The University of Texas at Austin, Austin, 1997.

[17] C.K. Birdsall and A.B. Langdon. Plasma physics via computer simulation. In *Lecture Notes*. University of California at Berkeley, 1980.

[18] J.A. Bittencourt. *Fundamentals of plasma physics*. Springer-Verlag, New York City, 3rd edition, 2004.

[19] J.P. Boris and D.L. Book. Solutions of continuity equations by the method of flux-corrected transport. *Journal of Computation Physics*, 20:397–431, 1976.

[20] S.C. Brenner and L.R. Scott. *The mathematical theory of finite element methods*. Springer-Verlag, New York City, 2002.

[21] F. Brezzi, G. Manzini, D. Marini, P. Pietra, and A. Russo. Discontinuous Galerkin approximations for elliptic problems. *Numerical Methods for Partial Differential Equations*, 16:365–378, 2000.

[22] J.A. Carillo, I.M. Gamba, A. Majorana, and C.W. Shu. 2D non-stationary Boltzmann-Poisson systems:DSMC0 versus a WENO-Boltzmann scheme. *Journal of Computational Physics*, 3:121–153, 2005.

[23] P. Castillo, B. Cockburn, D. Schötzau, and C. Schwab. An a priori error analysis of the local discontinuous Galerkin method for elliptic problems. *SIAM Journal on Numerical Analysis Numerical Analysis*, 38:1676–1706, 2000.

[24] C. Cercignani, R. Illner, and M. Pulvirenti. *The Boltzmann equation and its applications*. Springer-Verlag, New York City, 1st edition, 1987.

[25] C. Cercignani, R. Illner, and M. Pulvirenti. *The mathematical theory of dilute gases*. Springer-Verlag, New York City, 1st edition, 1994.

[26] F. F. Chen. *Introduction to plasma physics and controlled fusion*. Plenum Press, New York, 2nd edition, 1984.

[27] C. Cheng and G. Knorr. The integration of the Vlasov equation in configuration space. *Journal of Computational Physics*, 22:330–351, 1976.

[28] P.C. Clemmow and J.P. Dougherty. *Electrodynamics of particles and plasmas*. Addison-Wesley, Reading, 1969.

[29] B. Cockburn and C. Dawson. Some extensions of the local discontinuous Galerkin method for convection-diffussion equations. In *The Proceedings of the Conference on the Mathematics of Finite Elements and Applications*, pages 225–238. Elsevier, 2000.

[30] B. Cockburn, S. Hou, and C.W. Shu. The Runge-Kutta local projection discontinuous Galerkin method for conservation laws. IV: the multi-dimensional case. *Mathematics of Computation*, 54(190):545–581, 1990.

[31] B. Cockburn, S. Hou, and C.W. Shu. The Runge-Kutta local projection P1-discontinuous Galerkin method finite element method for scalar conservation laws. *Mathematical Modelling and Numerical Analysis*, 25:337–361, 1991.

[32] B. Cockburn, G.E. Karniadakis, and C.W. Shu, editors. *Lecture Notes in Computational Sciences and Engineering*. Springer-Verlag, New York, 2000.

[33] B. Cockburn, S. Y. Lin, and C.W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws iii: one dimensional systems. *Journal of Computational Physics*, 84:90–113, 1989.

[34] B. Cockburn and C.W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws ii: general framework. *Mathematics of Computation*, 52:411–435, 1989.

[35] B. Cockburn and C.W. Shu. The local discontinuous Galerkin method time-dependent

convection-diffusion systems. *SIAM Journal on Numerical Analysis Numerical Analysis*, 35(6):2440–2463, 1998.

[36] G.H. Cottet and P.A. Raviart. Particle methods for the one-dimensional Vlasov-Poisson equations. *SIAM Journal on Numerical Analysis*, 21:52–76, 1984.

[37] G.H. Cottet and P.A. Raviart. On particle-in-cell methods for the Vlasov-Poisson equations. *SIAM Journal on Numerical Analysis*, 26:1–31, 1986.

[38] C.N. Dawson, S. Sun, and M.F. Wheeler. Compatible algorithms for coupled flow and transport. *Computational Methods in Applied Mechanics and Engineering*, 193:2565–2580, 2004.

[39] J.M. Dawson. Plasma simulation of plasmas. *Reviews of Modern Physics*, 55:403–447, 1980.

[40] J. Douglas and T. Dupont. *Interior penalty procedures for elliptic and parabolic Galerkin methods*. Springer-Verlag, Berlin, 1976.

[41] J. Douglas, I.M. Gamba, and M.C.J. Squeff. Simulation of the transient behavior of a one dimensional semiconductor device. *Mat. Apl. Comput.*, 5:103–122, 1986.

[42] A. Ern and J.L. Guermond. *Theory and practice of finite elements*. Springer-Verlag, New York City, 2004.

[43] F. Filbet. Convergence of a finite volume scheme for the Vlasov-Poisson system. *SIAM Journal on Numerical Analysis*, 39(4):1146–1169, 2001.

[44] I.M. Gamba and M.C.J. Squeff. Simulation of the transient behavior of a one dimensional semiconductor device II. *SIAM Journal on Numerical Analysis*, 5:103–122, 1986.

[45] K. Ganguly and Jr. H.D. Victory. On the convergence of particle methods for multidimensional Vlasov-Poisson systems. *SIAM Journal on Numerical Analysis*, 26:249–288, 1989.

[46] K. Ganguly, J.T. Lee, and Jr. H.D. Victory. On simulation methods for Vlasov-Poisson systems with particles asymptotically distributed. *SIAM Journal on Numerical Analysis*, 28:1574–1609, 1991.

[47] R.J. Goldston and P.H. Rutherford. *Introduction to plasma physics*. Institute of Physics Publishing, London, 2nd edition, 1997.

[48] P. Grisvard. *Elliptic problems in nonsmooth domains*. Pitman Publishing, Boston, 1985.

[49] Jr. H.D. Victory, E.J. Allen, and K. Ganguly. The convergence theory of particle-in-cell methods for multidimensional Vlasov-Poisson systems. *SIAM Journal on Numerical Analysis*, 28:1207–1241, 1991.

[50] Jr. H.D. Victory, O. Tucker, and K. Ganguly. The convergence analysis of the fully discretized particle methods for solving Vlasov-Poisson systems. *SIAM Journal on Numerical Analysis*, 28:955–989, 1991.

[51] F. Herau and F. Nier. Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential. *Archive for Rational Mechanics and Analysis*, 171(2):151–218, 2004.

[52] R.W. Hockney and J.W. Eastwood. *Computer Simulation using Particles*. McGraw-Hill, New-York, 1981.

[53] H.J. Hwang. Regularity for the Vlasov-Poisson system in a convex domain. *SIAM Journal on Mathematical Analysis*, 36(1):121–171, 2004.

[54] G. Joyce, G. Knorr, and H. Meir. Numerical integration methods of the Vlasov equation. *Journal of Computational Physics*, 8:53–63, 1971.

[55] A.J. Klimas. A numerical method based on the fourier-fourier transform approach for modeling 1-d electron plasma evolution. *Journal of Computational Physics*, 50:270–306, 1983.

[56] A.J. Klimas and J. Cooper. VlasovMaxwell and VlasovPoisson equations as models of a one-dimensional electron plasma. *Physics of Fluids*, 26:478–479, 1983.

[57] A.J. Klimas and W.M. Farrell. A splitting algorithm for the Vlasov simulation with filamentation filtration. *Journal of Computational Physics*, 110:150–163, 1994.

[58] G. Knorr. Plasm simulation with few particles. *Journal of Computational Physics*, 13:165–180, 1973.

[59] L.D. Landau. On the vibrations of the electronic plasma. *J. Phys. U.S.S.R.*, 10(25):25–34, 1946.

[60] P. LeSaint and P.A. Raviart. On a finite element method for solving the neutron transport equation. *Mathematical aspects of finite elements in partial differential equations*, pages 89–123, 1974.

[61] P.L. Lions and B. Perthame. Propagation of moments and regularity of the 3-dimensional Vlasov-Poisson system. *Inventiones mathematicae*, 105:415–430, 1991.

[62] D.C. Montgomery and D.A. Tidman. *Plasma kinetic theory.* McGraw-Hill, New York, 1964.

[63] J.A. Nitsche. Über ein variationsprinzip zur Lösung von Dirichletproblemem bei Verwendung von Teilräumen, die keinen Randbedingun-gen unteworfen sind. *Anh. Math. Sem. Univ. Hamburg*, 36:9–15, 1971.

[64] J.T. Oden, I. Babuska, and C.E. Baumann. A discontinuous hp finite element method for diffusion problems. *Journal of Computational Physics*, 146:491–519, 1998.

[65] J.T. Oden and C.E. Baumann. A discontinuous *hp* finite element method for convection-diffusion problems. *Computational Methods in Applied Mechanics and Engineering*, 175(3-4):311–341, 1999.

[66] K. Pfaffelmoser. Global classical solutions of the Vlasov-Poisson system in three dimensions for general initial data. *Journal of Differential Equations*, 95:281–303, 1992.

[67] G. Rein. Collisionless kinetic equations from astrophysics - the Vlasov-Poisson-system. Work in progress.

[68] B. Riviere, M.F. Wheeler, and V. Girault. A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems. *SIAM Journal on Numerical Analysis Numerical Analysis*, 39(3):902–931, 2001.

[69] J. Schaeffer. Global existence of smooth solutions of the Vlasov-Poisson system in three dimensions. *Communications in Partial Differential Equations*, 16:1313–1335, 1991.

[70] H. Schamel. Electron holes, ion holes and double layers: electrostatic phase space structures in theory and experiment. In *Physics Reports (A Review Secton of Physics Letters)*, volume 140, pages 161–191. North-Holland, 1986.

[71] M. Shoucri and R.R.J. Gagne. Numerical solution of the Vlasov equation by transform methods. *Journal of Computational Physics*, 21:238–242, 1976.

[72] M. Shoucri and G. Knorr. Numerical integration of the Vlasov equation. *Journal of Computational Physics*, 14:84–92, 1974.

[73] E. Sonnendrücker, J. Roche, P. Bertrand, and A. Ghizzo. The semi-Lagrangian method for the numerical resolution of Vlasov equations. *Journal of Computation Physics*, 149:201–220, 1998.

[74] C. Villani and L. Desvillettes. On the trend to global equilibrium in spatially inhomogenous entropy-dissipating systems: the linear Fokker-Planck equation. *Communications on Pure and Applied Mathematics*, 54:1–42, 2001.

[75] M.F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM Journal on Numerical Analysis Numerical Analysis*, 15:152–161, 1978.

[76] M.F. Wheeler and P. Percell. A local residual finite element procedure for elliptic equations. *SIAM Journal on Numerical Analysis Numerical Analysis*, 15(5):705–714, 1978.

[77] S. Wollman. On the approximation of the Vlasov-Poisson system by particle methods. *SIAM Journal on Numerical Analysis Numerical Analysis*, 37:1369–1398, 2000.

[78] S. Wollman. The particle method for the Vlasov-Poisson system using equally spaced initial data points. *SIAM Journal on Computational and Applied Mathematics*, 115:593–600, 2000.

[79] S. Wollman, E. Ozizmir, and R. Narasimhan. The convergence of the particle method for the Vlasov-Poisson system with equally spaced initial data points. *Transport Theory and Statistical Physics*, 30(1):1–62, 2001.

[80] T. Zhou, Y. Guo, and C.W. Shu. Numerical study on Landau damping. *Physica D*, 157:322–333, 2001.

Vita

Ross Evan Heath, son of Rex Edward Heath and Rae Carol Heath, was born on May 26th in the year 1976, in Bellefontaine, Ohio. After graduating from Benjamin Logan High School, Bellefontaine, Ohio in 1994, he enrolled at The Ohio State University. In May 1999, he received the degree of Bachelor of Science in applied mathematics. He then went on to receive the degree of Master of Science in financial mathematics at The University of Chicago, in June 2000. After spending a year studying at The Weizmann Institute of Science, he began his graduate work at The University of Texas at Austin in August 2001.

Permanent Address: 12345 Lamplight Village Ave., Apt.1213
Austin, TX 78758

This dissertation was typeset with LaTeX 2_ε[1] by the author.

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin, and extended by Bert Kay, James A. Bednar, and Ayman El-Khashab.